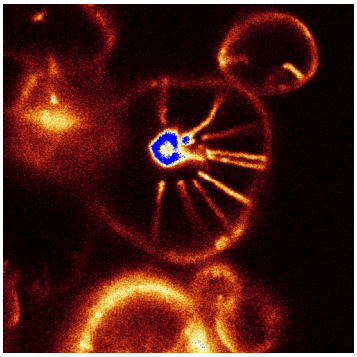
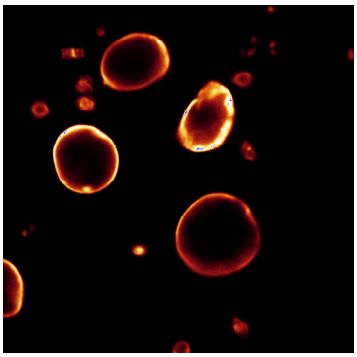
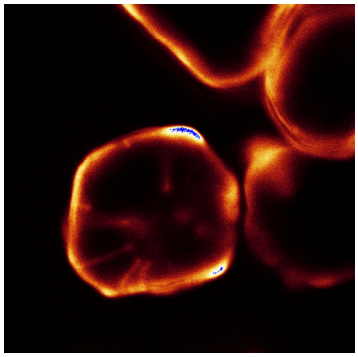
Investigation of Starch Binding Domains for Improvement of Starch degradation

Christiansen, Camilla; Blennow, Anders; Viksø-Nielsen, Anders; Svensson, Birte

Publication date:
2009

Document Version
Publisher's PDF, also known as Version of record

Investigation of Starch Binding Domains for Improvement of Starch degradation

Camilla Christiansen

Ph.D. Thesis

February 2009

Molecular Plant Biology
Department of Plant Biology and Biotechnology
Faculty of Life Sciences, University of Copenhagen

FOBI forskerskole for biofeknologi

DTU

novozymes Rethink Tomorrow

Investigation of Starch Binding Domains for Improvement of Starch degradation

Camilla Christiansen
Ph.D. Thesis

February 2009

Molecular Plant Biology
Department of Plant Biology and Biotechnology
Faculty of Life Sciences, University of Copenhagen

Preface

The present Ph.D. thesis entitled "Investigation of starch binding domain for improvement of starch degradation" summarises my work carried out at Department of Plant Biology and Biotechnology, Faculty of Life Sciences at University of Copenhagen and at Department of System Biology, Enzyme and Protein Chemistry, Technical University of Denmark.
The studies were financially supported by the research school for biotechnology (FOBI), Faculty of Life Sciences at University of Copenhagen and Novozymes.

I would like to acknowledge my supervisors; Associate Professor Andreas Blennow and Associate Professor Lone Baunsgaard from University of Copenhagen, Professor Birte Svensson, Associate Professor Henrik Næsted and Associate Professor Maher Abou Hachem from the Technical University of Denmark, Department of Systems Biology, Science manager Anders Viksø-Nielsen and Professor Lene Lange from Novozymes.

I would like to acknowledge all members – former and present – of the Plant Biochemistry Laboratory at at University of Copenhagen and Enzyme and Protein Chemistry at the Department of Systems Biology for providing a good working environment. A special thanks to Malene, Maher, Eun-Seong and Sabrina for their good spirits. Further more I would like to thank Stefan Janecek, Esben Friis, Mikkel Glaring, Lis Byrsting Møller, Morten Munch Nielsen and Martin J. Bauman for help and good collaboration during these last years and Jørgen Sauer for the original idea of fluorescence labelling the proteins.

Finally my friends and family for their support during the whole period

Camilla Christiansen

1 Table of contents

2 Summary

Starch is the primary energy storage of higher plants and an important nutrient for mammals, fungi and bacteria. Starch is deposited as highly organised semi-crystalline granules in plastids: chloroplasts in leaves (transitory starch) and amyloplasts in storage organs. Starch is entirely composed of α-1,4-linked glucose residues organised into the essentially linear amylose and the branched amylopectin containing α-1,6-linkages. Root and tuber starches are characterized by a high content of starch-bound phosphate. The presence of starch-bound phosphate has profound effects on both the physical properties of starch and its degradation in the plant. For industrial purposes the inclusion of phosphate esters, either naturally or by chemical modification, influences a number of very desirable properties. The enzyme responsible for incorporation of phosphate groups has previously been identified as an α-glucan, water dikinase (GWD). GWD phosphorylates the C-3 and C-6 positions of the glucose residues of starch in a dikinase-type reaction using the β-phosphate of ATP. Mutants in *Arabidopsis thaliana GWD1* display a starch-excess phenotype with reduced rates of starch degradation, indicating a link between starch phosphorylation and degradation.

Two homologues of *AtGWD1* were identified in the *Arabidopsis* genome and named *AtGWD2* and *AtGWD3*. Also mutations in *AtGWD3* resulted in a starch-excess phenotype, suggesting an involvement in starch degradation. *At*GWD3 was shown to be chloroplastic and substrate analysis of purified *At*GWD3 demonstrated a distinct preference for prephosphorylated α-glucans and catalysed exclusive phosphorylation at the C-3 position. Taken together, these data suggests that *At*GWD1 and *At*GWD3 act together in a sequential phosphorylation cascade that ensures complete breakdown of starch. The degradation of starch by glycoside hydrolases is relatively inefficient as the polysaccharide chains often are barely inaccessible to the active site of the enzymes. Many raw starch-degrading enzymes contain secondary binding sites on the catalytic domain or on separate starch binding domains (SBD) which facilitates the interaction. SBDs are classified according to the CAZy classification system in different carbohydrate binding module (CBM) families. *At*GWD1 and *At*GWD2 have a double repeat of an SBD belonging to family CBM45 and *At*GWD3 has a single SBD from family CBM20, all are N-terminally placed (Paper IV).

The aim of the present Ph.D. project has been to investigate and characterise the biochemical function of the GWD3-SBD. The GWD3-SBD has been successfully expressed as an isolated domain in *E. coli*. and purified. Stability of the domain increased as additional amino acids have been included at the domain border. Binding to raw starch has been difficult to measure, and this is most likely a consequence of the highly specialised role that *At*GWD3 plays in the starch phosphorylation. The enzyme is regulated in the chloroplast and only very specific areas of the starch granule are suited for glucan phosphorylation in the C3 position and finding optimal binding conditions has been difficult. The binding constant to small oligosaccharides like β-cyclodextrins (β-CD) was determined by Surface plasmon resonance and the domain is described as a low affinity binder, a category that includes other non-amylolytic enzymes (Paper I, III). The overall CBM20 fold

is conserved in the GWD3-SBD based on a homology model and at least binding site 1, which is involved in initial binding and conserved at the structural and the sequence level. Compared to other characterised CBM20 the GWD3-SBD sequence has a smaller loop in binding site 2 and the flexibility of this loop seen upon ligand binding in other CBM20s might explain the lower binding capacity of the domain. Fluorescent labelling and the usage of confocal laser scanning microscopy has been applied as a method to visualise binding of the SBDs and hydrolytic enzymes, i.e. fungal glucoamylase and barley α-amylase to starch granules (Paper II). This method was used together with transient expression in tobacco of a yellow-fluorescent protein YFP-GWD3-SBD fusion to verify starch binding.

3 Dansk resume

Stivelse er planternes primære energilager og et vigtigt næringsmiddel for pattedyr, svampe og bakterier. Stivelse deponeres i højt organiserede semi-krystallinske stivelseskorn i plastider: kloroplaster i blade (transitorisk stivelse) og amyloplaster i lagerorganer som knolde. Stivelse består udelukkende af α-1,4-bundne glukose enheder, som er organiseret enten som det stort set lineære amylose molekyle eller det forgrenede amylopektin molekyle, der indeholder α-1,6-bindinger. Rod og knold stivelse er karakteriseret ved et højt niveau af kovalent bundet fosfat. Denne stivelsesbundne fosfat har en stor effekt på både stivelsens fysiske egenskaber samt på stivelsesnedbrydning i planterne. Inkludering af fosfatestre påvirker de industrielle egenskaber, og medfører forskellige meget ønskværdige egenskaber. Enzymet der kan inkorporere fosfat grupper i stivelse er en glucan, water dikinase (GWD). GWD kan fosforylere i C-3 og C-6 positionen i glukose enhederne i stivelse, ved en dikinase reaktion der anvender β-fosfat fra ATP. Mutanter i *Arabidopsis thaliana GWD1* udviser en stivelses overskud fænotype med en lavere stivelses nedbrydnings rate, hvilket påviser en forbindelse mellem stivelses fosforylering og stivelses nedbrydning.

To homologe proteiner er blevet identificeret i *Arabidopsis* genomet, navngivet *At*GWD2 og *At*GWD3. Mutationer i *At*GWD3 resulterede også i en stivelses overskud fænotype, som set hos *At*GWD1, hvilket tyder på *At*GWD3 også er involveret i stivelses nedbrydning. *At*GWD3 er lokaliseret i kloroplasterne og substrat analyser viste at oprenset *At*GWD3 udviser præference for fosforyleret α-glucaner og katalysere udelukkende fosforylering i C-3 positionen i glukose enhederne. Disse resultater tyder på at *At*GWD1 og *At*GWD3 arbejder sammen i en efterfølgende fosforyleringskaskade, som er nødvendig for nedbrydning af stivelse.

Glycosyl hydrolasers nedbrydning af rå stivelse er relativ ineffektiv, da polysacharrid kæderne ofte ikke er blotlagte og tilgængelige for enzymernes aktive site. Mange stivelses nedbrydende enzymer har ekstra bindings sites i det katalytiske domæne eller på separate stivelsesbindings domæner (SBD) som muliggør denne interaktion. SBDer er klassificeret i CAZy databasen i forskellige kulhydrat bindings module (CBM) familier. *At*GWD1 og *At*GWD2 har et tandem repeat af SBDer som tilhører familie CBM45 og *At*GWD3 har et enkelt SBD fra familie CBM20. Alle er N-terminalt lokaliseret (Publikation IV).

Formålet med dette PhD projekt har været at undersøge og karakterisere GWD3-SBDs biokemiske funktion. GWD3-SBD er blevet udtrykt succesfuldt som et isoleret domæne i *E. coli* og oprenset. Øget stabilitet af domænet blev opnået efter

yderligere aminosyrer blev inddraget i den kodende sekvens. Binding til stivelseskorn var svært at måle, og det skyldes højst sandsynligt konsekvensen af den højt specialiserede rolle som GWD3 spiller i stivelses fosforyleringen. Enzymet er reguleret i kloroplasten og kun meget specialiserede områder på stivelses kornene er egnede for fosforylering i C-3 positionen og det er svært at finde forhold, hvor alle parametre er optimale. Binding til små oligosaccharider, som β-cyclodextrins (β-CD) er blevet bestemt med Surface plasmon resonance og domænet tilhører en gruppe af lav affinitets bindere, denne kategori inkluderer ikke-hydrolyserende enzymer (Publikation I, III). Den overordnede struktur fundet hos CBM20 er ifølge en homologimodellering bevaret i GWD3-SBD og bindings site 1, som er involveret i initial binding er vel bevaret både i strukturen og på sekvens niveau. Sammenlignet med andre karakteriserede CBM20, så har GWD3-SBD et mindre loop i området omkring bindings site 2. Dette loop er under substrat binding i andre CBM20 meget fleksibelt og dette kan forklare den lavere bindings kapacitet fundet hos GWD3-SBD. Fluorescens mærkning og confocal laser scanning microskopi er blevet anvendt som en metode til at visualisere SBD og hydrolyserende enzymers, f.eks. glucoamylase og α-amylases binding til stivelseskorn (Publikation II). Denne metode blev anvendt sammen med transient ekspression af en yellow-fluorescent protein YFP-GWD3-SBD fusion i tobaksplanter for at verificeres stivelsesbinding .

4 Abbreviations

ADPG	ADP-glucose
AGPase	ADP-glucose pyrophosphorylase
AMY	α-amylase
ATP	Adenosine triphosphate
BAM	β-amylase
CBM	Carbohydrate binding module
CD	Cyclodextrin
CLSM	Confocal laser scanning microscopy
CV	Column volume
DBE	Debranching enzyme
DTT	Dithiothreitol
DP	Degree of polymerization
DPE	Disproportionating enzyme
Fru6P	Fructose-6-phosphate
GA	Glucoamylase
GBSS	Granule-bound starch synthase
GH	Glycoside hydrolase
Glc1P	Glucose-1-phosphate
Glc6P	Glucose-6-phosphate
GST	Glutathione S-transferase
GUS	β-glucoronidase
GWD	α-gucan, water dikinase
ISA	Isoamylase
ITC	Isothermal titration calorimetry
LDA	Limit dextrinase
MOS	Malto-oligosaccharides
PGI	Phosphoglucoisomerase
PGM	Phosphoglucomutase
PHS	α-gucan phosphorylase
PPDK	Pyruvate phosphate dikinase
SBD	Starch binding domain
SBE	Starch branching enzyme
SDS-PAGE	Sodium dodecyl sulphate-polyacrylamide gel electrophoresis
SEX	Starch excess phenotype
SPR	Surface plasmon resonance
SS	Starch synthase
RSMD	Root mean square
YFP	Yellow fluorescent protein

5 List of publications

I Christiansen, C, Abou Hachem, M., Friis, E., Baumann, M.E., Glaring, M.A., Viksø-Nielsen, A., Sigurskjold, B.W., Svensson, B. and Blennow, A. (2008) Exploring New Plant Carbohydrate Binding Modules (CBMs) from Glucan, Water Dikinase. In: *Starch Recent Progress in Biopolymer and Enzyme Technology*. Tomasik, P., Bertoft, E. and Blennow A. (Editors). Polish Society of Food Technologists, Chapter 7, pp. 85-102

II Seo, E.S., Christiansen, C., Abou Hachem, M., Nielsen, M.M., Fukuda, K., Bozonnet, S., Blennow, A., Aghajari, N., Haser R. and Svensson, B. (2008) An enzyme family reunion - similarities, differences and eccentricities in actions on α-glucans. *Biologia, Bratislava* Volume 63, Issue No. 6, pp 967-979.

III Christiansen, C., Abou Hachem, M., Glaring, M.A., Viksø-Nielsen, A., Sigurskjold, B.W., Svensson, B. and Blennow, A. (2009) A CBM20 low-affinity starch binding domain from glucan, water dikinase. Submitted to *FEBS letter*.

IV Christiansen, C., Abou Hachem, M., Janeček, Š, Viksø-Nielsen, A., Svensson, B. and Blennow, A. (2009). The carbohydrate binding module family 20 diversity, structure and function. *Draft of Review manuscript.*

6 Introduction

The aim of the present Ph.D. thesis is to characterise the function of starch binding domains (SBDs) from carbohydrate binding module family 20 (family CBM20) and explore how they are involved in starch degradation. The efficiency of starch degradation in both the plants and in starch processing units is a delicate balance between several parameters including the enzyme intrinsic catalytic activity and specificity, its starch-binding capacity and affinity as well as the starch type. To understand the mechanism behind degradation and mobilisation of starch a thorough knowledge of starch granule structure and organisation is needed. The introduction in this thesis will provide information on starch granule structure, biosynthesis and degradation of starch. The starch phosphorylating enzyme α-glucan, water dikinase 3 (GWD3) possesses a CBM20 and this CBM has been the focus and plays a key role throughout this thesis. SBDs are present in many structurally and functionally different enzymes involved in starch breakdown. A lack of reviews in the scientific literature describing the structure-function relationships in the CBM20 family prompted us to make an update of recent advances on this topic (Paper IV).

The following questions initiated the current study: What role is the CBM20 in the plant GWD3 enzyme playing with regard to attacking and initialising the degradation or modification of starch granules? Can these domains function as isolated modules? Can we identify any substrate preference? Since this domain is the first reported of an SBD of plant origin to be studied and since it is present in a regulatory enzyme, what possible specific differences or similarities does it have to other CBM20 members? Is the starch affinity depending on the starch granule type and can it be investigated *in vitro* also in the form of a separate module?

Understanding starch binding affinity and how to alter this by engineering SBDs and by modifying already existing hydrolytic enzymes by adding or exchanging SBDs, is a very interesting area in relation to industrial processing of starch. To further address the function of the GWD3-SBD and analyse the effect of exchanging the SBD of existing hydrolytic enzyme by GWD3-SBD, a fusion protein was produced by replacing the *A. niger* glucoamylase SBD by the GWD3-SBD. Also binding mechanisms of glucoamylase and barley α-amylase to raw starch have been investigated in this thesis.

The project is important in focusing on a poorly studied and characterised category of putative SBDs present in non-amylolytic enzymes. In addition, it will increase our basic knowledge of SBD specificity and binding efficiency for selected important starch types. The outcome is significant for a better understanding of the regulation of starch metabolism and degradation and the development of novel efficient and sustainable enzyme catalysing starch modification and liquefaction processes.

6.1 Starch

Starch is a very important storage polysaccharide in all plants and is the most abundant food biopolymer worldwide. It constitutes the primary intake of calories in the human diet (Myers *et al.,* 2000). Starch is a cheap and easily accessible raw material which is exploited in the agricultural food sector. It occurs in a variety of botanical sources including potato, wheat, cassava and maize and has found diverse applications ranging from breakfast cereals, snacks and thickeners in the food industry to textiles, paper, binders for drug delivery systems and adhesives in the non-food industry. Most starch used in industry is degraded to generate glucose and high fructose syrups used as sweeteners (approx 60%) and the remainder is modified to generate highly different desired functionalities (http://www.starch.dk). Starch is deposited in the plastids as granules which differ in shape and size for different plant species (Gallant *et al.*, 1997). The plastids can be photosynthetic *i.e.* chloroplasts synthesizing transient leaf starch or non-photosynthetic such as amyloplasts where storage starch is synthesised and deposited. Storage starch is found in plant organs like seeds, fruits, tubers and roots and functions as a source of energy reserve to initialise re-growth after dormancy periods and has a lower turnover than chloroplast starch (Jobling, 2004). Variations in the starch granules architecture and physiochemical characteristics between different plants species affect the properties and functions of starches (Jane *et al.*, 1994). Starch is a relatively simple polymer composed exclusively of α-D-glucose residues linked together in two different molecules, amylose and amylopectin (Fig. 1).

6.1.1 Amylose

Amylose is a mainly linear molecule in which the glucose units are linked by α-1,4 glucosidic bonds. Some amylose molecules contain α-1,6 linked branch points at very low frequencies, but this does not significantly alter the behaviour of the amylose chains in solution (Takeda and Hizukuri, 1987; Buléon *et al.,* 1998). Linear regions of the amylose chains form a dark blue complex with polyiodine in aqueous solution at room temperature (Yu *et al.*, 1996). Amylose binds 20% of its weight of iodine while under the same conditions amylopectin generally binds < 1% w/w. This iodine binding allows a distinction to be made between amylose and amylopectin and permits determination of the amylose content of native starch. The amylose content may vary with the plant species, and constitutes 20-25% by weight and furthermore the value can be highly variable depending on the developmental stages of the plant organ. The molecular weight of amylose is in the range of 10^5-10^6 Da with a degree of polymerization (DP) ranging from a few hundred to thousands, but amylose is still about 100 times smaller than amylopectin.

6.1.2 Amylopectin

The major component of starch, amylopectin is considered to be one of the largest naturally occurring macromolecules with a molecular weight of 10^7-10^9 Da. Amylopectin is a branched polysaccharide composed of hundreds of short (1→4) α-glucan chains, which are interlinked by (1→6) α-linkages (Fig. 1). The α-1,6 linked branch points constitutes about 5% of the linkages in the amylopectin molecule.

The branching within amylopectin is not random and occurs at regular intervals, resulting in a clustering of a large number of shorter linear chains (Hizukuri, 1986; Manners, 1989). Debranching amylopectin with a debranching enzyme such as isoamylase followed by size exclusion chromatography reveals an essential bimodal population of chains with two main populations with peak DP of around 12-14 and 45, respectively (Hizukuri, 1985).

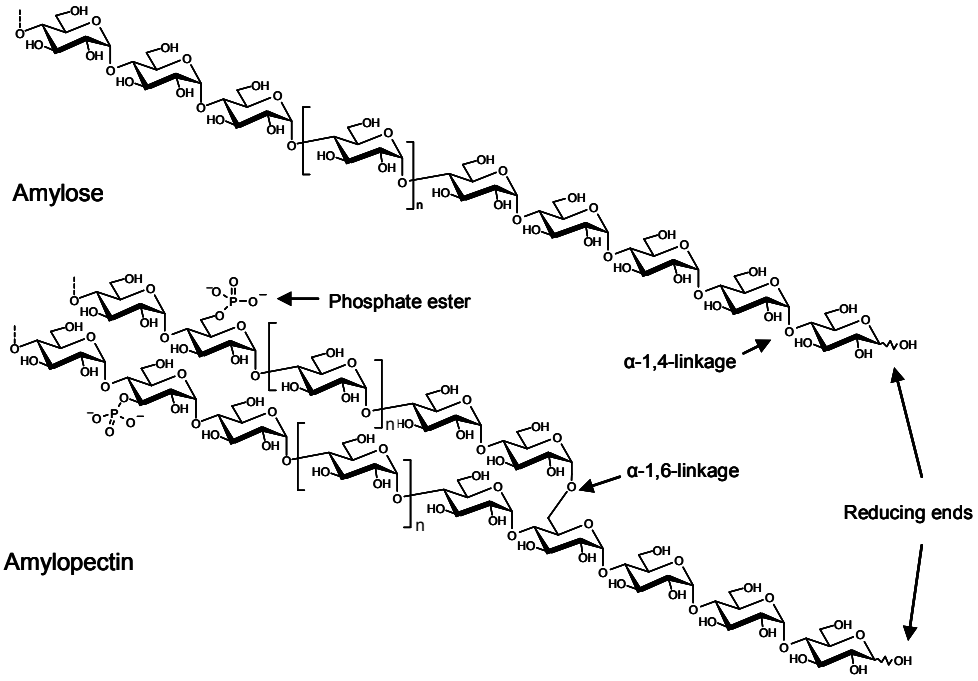


Figure 1. Structure of amylose and amylopectin.
Amylose is a polysaccharide composed of glucose residues joined by α-1,4 linkages. Amylopectin is a polymer glucose with an α-1,4 backbone and an additional 5% α-1,6 linkages (branch points). The only naturally occurring modification: monoesterified phosphate is shown on the glucose residues (C-3 or C-6 position): n indicates the variable length of the glucan chains.

The short chain fractions of many amylopectins also have shoulders of around DP 18-20 (Hanashiro *et al.*, 1996). The proportion of short to long chains depends on the plant species and varies typically from roughly 5 ~ 13 on a molar basis (Hanashiro *et al.,* 2002). The organisation of the chains can be described in terms of A, B and C chains (Fig. 2). A-chains are linked only through their reducing end and contain therefore no branch points. The B-chains are attached to an inner chain and are carry multiple chains. B chains are divided according to the length distribution where B1 chains are suggested to span only one cluster whereas B2 and B3 etc. are named according to the number of clusters in which they participate. The single C chain per molecule contains the reducing end, O in the polysaccharide (Fig. 2). Glycogen, displays the same basic structural features as amylopectin, but has asymmetrical distribution of unit chains of average length 10 to 14 glucose residues, and cannot form the long-range solid-state ordering as found in starch (Manners, 1991).

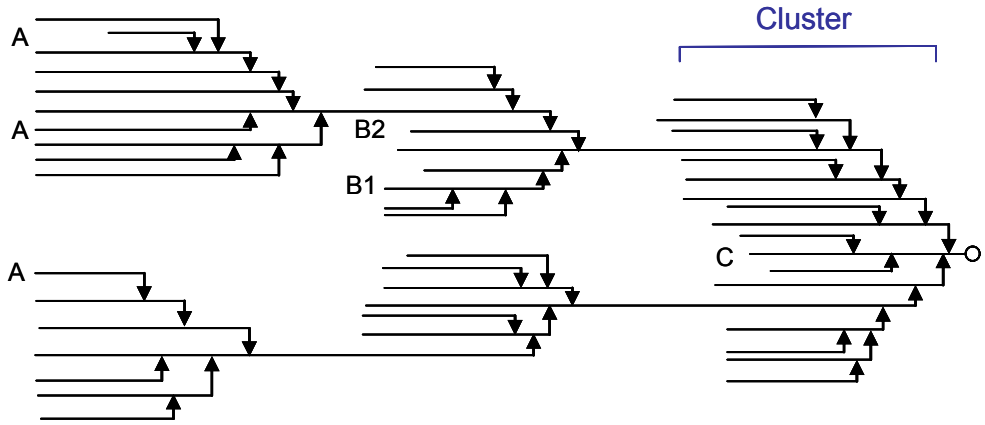


Figure 2. Cluster model of amylopectin structure.
The cluster model of showing A, B1, B2 and C chains (Hizukuri, 1986). The C chain carries only the reducing end, O, in the polysaccharide.

6.1.3 Starch granule organisation

In the plant, starch is deposited as water insoluble semi-crystalline granules that vary from less than 1 μm to over 100 μm in size. Many granules approximate to spheres, but some are more elongated (reviewed by (Smith, 2001). Some cereals including oat, wheat and barley has a dual population of granules with both large (A-granules) and small (B-granules), whose biosynthesis occur at two different stages of development (Evers, 1971). In wheat, A-granules are between 10 and 38 μm in size and disc- or lenticular-shaped, while B-granules are smaller than 10 μm and possess a spherical or polygonal morphology (Wilson *et al.*, 2006). Beside morphological differences, wheat starch A- and B-type granules also possess different characteristics and properties with regard to chemical composition (amylose, amylose–lipid complex, and phosphorus contents) (Geera *et al.,* 2006; Ao and Jane, 2007).

The point of initiation of the granule is called the hilum. It can be found near the centre of elliptical granules or positioned on the axis of symmetry, at the fat end of pear-shaped granules. It is usually considered to be less organised than the rest of the granule and there is general agreement that the chains within the granule are radially arranged with their nonreducing ends pointing toward the surface (French, 1973). This arrangement is suggested from the positive birefringence found for most starch granule types (Blennow *et al.*, 2003).

The presence of branch point at the root of each cluster allows the short linear glucans to form parallel double helices. The parallel packing of these double helices are responsible for the crystallinity of starch and can be monitored by X-ray diffraction analysis. The crystalline lamellae alternate with amorphous lamellae formed by the regions in which the branch points occur with a periodicity of 9 nm. The repeat distance of the lamellae is highly conserved in the plant kingdom (Jenkins *et al.,* 1993) suggesting that starch biosynthesis is highly conserved. The alternating

lamellae form concentric, semicrystalline zones within the granule and with a periodicity of a few hundreds nanometers (Fig. 3). Between the semicrystalline zones, there are amorphous zones, in which amylopectin molecules are in a less organised state, and in which much of the amylose component of the granule is thought to occur (Jenkins *et al.,* 1993; Waigh *et al.,* 1998). One repeat on this scale is known as a growth ring. In general, it seems that amylose is not necessary for obtaining a classic semi-crystalline starch structure as observed by amylose free *waxy* cereals as well as pea (reviewed by (Smith *et al.,* 1995; Buléon *et al.,* 1998; Smith, 2001).

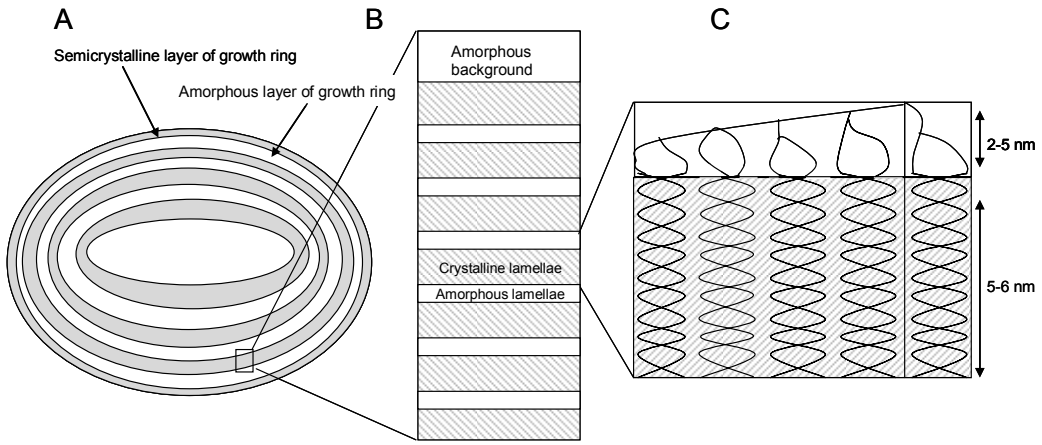


Figure 3. An overview of starch granule organisation.
A: Schematic view of a starch granule with its succession of amorphous and crystalline growth rings. **B**: A section of a crystalline growth ring of the granule related to the molecular organisation of amylopectin. Each shaded and plain section represents an amorphous and a crystalline lamella respectively. Thus the crystalline growth ring enlarged in this panel contains a regular succession of 10 amorphous and crystalline lamellae. This would amount to a 0.1 µm thick growth ring. **C**: This panel show the secondary structure of a single cluster. Within a cluster, adjacent chains form double helices. These pack together in a regular manner to form crystalline lamellae, which alternate with amorphous lamellae where the branch points are located Modified from Ball and Morell, (2003).

However amylose affects the crystalline arrays of the starch granule by modulating the lamellar thickness and by introducing crystalline defects at high concentrations (Kozlov *et al.*, 2007). The original cluster model presented in Figure 2 has been updated with some later findings, which relates the model to the semicrystalline structure as it appears inside the starch granules. A new two-directional backbone model have been proposed (Bertoft, 2004) in which the long B chains are removed from their positions as integrated parts of the clusters and have their directions perpendicular to that of the internal clustered chains and the long B-chains are placed into the amorphous lamella. A large amount of long chains compared to the number of clusters, and a large number of glucosyl residues found at inter-cluster areas was found in a recent investigation of the cluster structure of amylose-free potato starch (Bertoft, 2007b). Further investigation of the amorphous lamellae showed that the chains of the clusters of potato amylopectin were organised into densely branched

building blocks interconnected by internal segments of 7–8 glucosyl residues (Bertoft, 2007a). A super helix model in which the amylopectin is arranged into a continuous network of left-handed super helices with a diameter of 18 nm and a pitch of 10 nm has been proposed (Oostergetel and van Bruggen, 1993). The two-directional backbone model can also be adapted to a super-helix and the entire super helix is build of a single amylopectin molecule (Bertoft, 2004).
Native starch exhibits two types of X-ray diffraction patterns that are associated to two polymorph forms: the A-type found mainly in cereal endosperm starches and the B-type found mainly in tuber starches (Imberty, 1988; Imberty *et al.,* 1988) (Fig. 4). In crystallites of both A- and B- type starch, double helices are found in pairs, and all chains are packed in parallel arrays. The pairing of double helices is identical in both polymorphs and corresponds to the interaction between double helices that have the lowest energy. The difference between A and B starch arises from water content and the manner in which these pairs are packed in respective crystals. The A-type is made from left-handed parallel double helices packed in a dense structure with 8 water molecules per unit cell and contains more abundant and shorter A and B1 chains (Imberty *et al.,* 1988; Hanashiro *et al.,* 2002). The B-type is also packed with left-handed parallel double helices but has an open structure with 36 water molecules per unit.

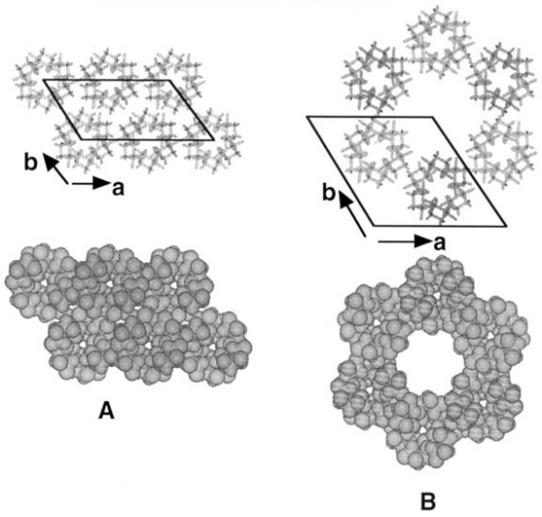


Figure 4. A and B type starch
Double helices can be arranged in the A-type, B-type or an intermediary C-type. The B structure display a more hydrated capacity than the A structure. Modified from (Buléon *et al.,* 1998).

Pores have been observed on the surfaces of some corn, sorghum, millet granules and along the equatorial groove of large granules of wheat, rye and barley (Buléon *et al.,* 1998). Their distribution appeared to be random, with a variable numbers per granule (Huber and BeMiller, 2000). The existence of central cavities in starch granules from different botanical origins has been demonstrated and that the cavities are connected to the granule exterior by a network of channels extending to the granule surface (Huber and BeMiller, 2000; Glaring *et al.,* 2006). Patterns of

enzymatic attack of wheat starch have been observed to produce large pin holes at granule surfaces (along the equatorial groove).
Besides amylose and amylopectin a number of minor components such as lipids, proteins and phosphate influence the physical and chemical properties of starch. Lipids, which are embedded in the granule, represent the most important of these fractions. Cereal starches generally contain the highest lipid content about 0.6-1.2% (Buléon *et al.,* 1998). The amount of lipid is positively correlated to the amylose content, presumably as a consequence of their ability to form complexes with amylose. Triglycerides represent a major fraction of surface lipids of maize and wheat starch (Vasanthan and Hoover, 1992). Proteins found within the starch granule are proteins directly involved in the biosynthesis of starch, like the amylose synthesising granule bound starch synthase 1 (GBSS1).

6.1.4 Covalently bound phosphate

A small fraction of the glucose units (0.1-1%) in amylopectin contains covalently bound phosphate. Phosphate groups are bound as monoesters at the C3 or C6 position (Fig. 1). Approximately 60-70% of the phosphate groups are bound as monoesters to the C-6 positions of the glucosyl units, whereas the remaining 30-40% are monoesterified to the C-3 position (Tabata and Hizukuri, 1971). Phosphate groups are mainly localised in the longer unit chains of DP 30-100 (Blennow *et al.,* 1998; Blennow *et al.,* 2002) and no phosphate groups are located on the non reducing end or closer than 10 glucosyl units from an α-1,6 branch point (Takeda and Hizukuri, 1981). The chain length distribution of amylopectin shows a positive correlation between the DP of un-substituted chains and the degree of phosphorylation (Blennow *et al.,* 1998). The phosphate monoesters are incorporated into starch during biosynthesis (Nielsen *et al.,* 1994) and have profound effects on both the physical properties of starch and its degradation in the plant (Blennow *et al.,* 2002). The glucan, water dikinase responsible for phosphorylation will be described in detail later.
Endosperm starches contain little or no covalently starch-bound phosphate (<0.04% w/w) and tuberous starches tend to have the highest contents of starch phosphate (0.2-0.4% w/w) (Blennow *et al.,* 2000). However, leaf starch also contains decent amounts of phosphate (Yu *et al.,* 2001) so endosperm starch seems to be the exception. Both crystalline and amorphous amylopectin regions contain esterified phosphate (Blennow *et al.,* 2000). Molecular modelling of amylopectin double-helical motifs has indicated that C-6-bound phosphate groups align well in the double helix surface grooves, and when in this position, the phosphate will not affect the double helix formation severely. By contrast, the C-3-bound phosphate will protrude from the double-helical surface and have a greater effect on starch double helix packing (Engelsen *et al.*, 2003). The presence of covalently bound phosphate groups in many starches, has a major impact on the rheological properties of isolated starch, causing clearer gels and higher viscosity, which is advantageous for many industrial applications (Jobling, 2004).

6.1.5 Starch properties

Starch granules are insoluble in cold water and before gelatinization (loss of crystallinity) occurs, the granules require heating to approximately 60 ˚C for most normal starch types. The double helices within the amylopectin molecule melt upon heating in water, hereby the starch granule starts to swell and the viscosity increases in the solution. Cooling of the solution re-associates the linear chains into aggregates which can precipitate into an insoluble form, *i.e.* so called retrogradation (French, 1973; Parker and Ring, 2001). The rheological properties also greatly depend on the starch molecular weight distribution (Blennow *et al.*, 2001). Quite contrary to amylose, amylopectin has more limited hydrogen bonding in solution and is more stable. It remains fluid and gives high viscosity and elasticity to pastes and thickeners. The phosphorylation of starch contributes to a range of physico-chemical properties such as pasting properties, gel strength and stickiness and it is necessary to prevent crystallization of the final product when utilised in industrial products (Wiesenborn *et al.,* 1994; Viksø-Nielsen *et al.,* 2001; Nabeshima and Grossmann, 2001). The effect of starch phosphorylation in the potato tuber was found to increase the hydration capacity of starch pastes providing clearer gels and better freeze-thaw stability (Viksø-Nielsen *et al.*, 2001).

Novel starches with improved functionality are now being produced using genetic modification to alter expression of individual starch biosynthetic genes. Mutation in the *Waxy* locus, which encodes the GBSS protein, creates a starch type without amylose (Shure *et al.,* 1983). The physical properties of waxy starches include rapid gelatinization, a clearer paste that does not gel as well as improved freeze-thaw stability as compared with normal starches (Zheng and Sosulski, 1998; Jobling *et al.,* 2002). On the other hand, high amylose starches are interesting due to their high gelling strength and have been produced by manipulating the degree of branching suppressing starch branching enzyme activity (Schwall *et al.*, 2000). High-amylose starches can be processed into resistant starch, which has nutritional benefits (Bird *et al.*, 2000). Unlike normal starch, resistant starch is not digested in the small intestine but fermented in the large intestine by gut bacteria, producing short-chain fatty acids that are beneficial for colon health.

6.2 Biosynthesis of starch

During photosynthesis, energy from solar radiation is used for the formation of phosphorylated C3 sugar phosphates, and these products are exported from the chloroplasts into the cytosol via a translocator. In the mature leaves of most plants, the exported photosynthates are then used in the formation of sucrose, which is allocated via the phloem to the starch-storing organs such as potato tubers and endosperm of cereal grains. Sucrose serves as a source of carbon and after cleavage by invertases or sucrose synthase hexose phosphates can enter the amyloplast via a transporter in the membrane (Kammerer *et al.*, 1998). The overall biosynthesis of starch can be divided into the following three steps which are valid for higher plants as well as algae:

(1) ATP + α-glucose-1-P ↔ ADP-G + PPi
(2) ADP-G + α-1,4-glucose → ADP + α-1,4-glucosyl-α-1,4-glucan
(3) α-1,4-glucan oligosaccharide → branched α-1,4-/α-1,6-glucan

In the first step ADP-glucose pyrophosphorylase (AGPase) catalyses the conversion of glucose-1-phosphate and ATP to the charged sugar nucleotide ADP-glucose and pyrophosphate. The action of AGPase represents the first committed step in starch biosynthesis. The enzyme is regarded as the major regulator of starch biosynthesis because its activity is influenced by several important metabolites. AGPase is in most cases allosterically regulated by the levels of 3-phosphoglycerate (3-PGA) generated from the Calvin cycle and phosphate (reviewed by (Ballicora *et al.*, 2004)). 3-PGA indicates carbon excess and activates AGPase, availability of ATP is modulated by the presence of AMP, ADP and Pi which signals for inhibition of ADP-glucose synthesis.
In the next step (2) starch synthases (SS) add glucose residues from ADP-glucose to the non-reducing end of a growing glucan chain. Multiple isoforms of SS have been found in higher plants (Marshall *et al.,* 1996). One group of SS genes contain the granule-bound starch synthases (GBSS), and includes GBSSI and GBSSII. GBSSI encoded by the *Waxy* locus in cereals, functions specifically to elongate amylose (Shure *et al.,* 1983). Mutations in the *Waxy* locus leading to loss of function of GBSS activity resulting in lack of amylose giving starches comprised solely of amylopectin. Similar mutants have been obtained in many other species including rice, barley, wheat, pea, potato, and *Chlamydomonas* (reviewed in (Ball *et al.*, 1998).

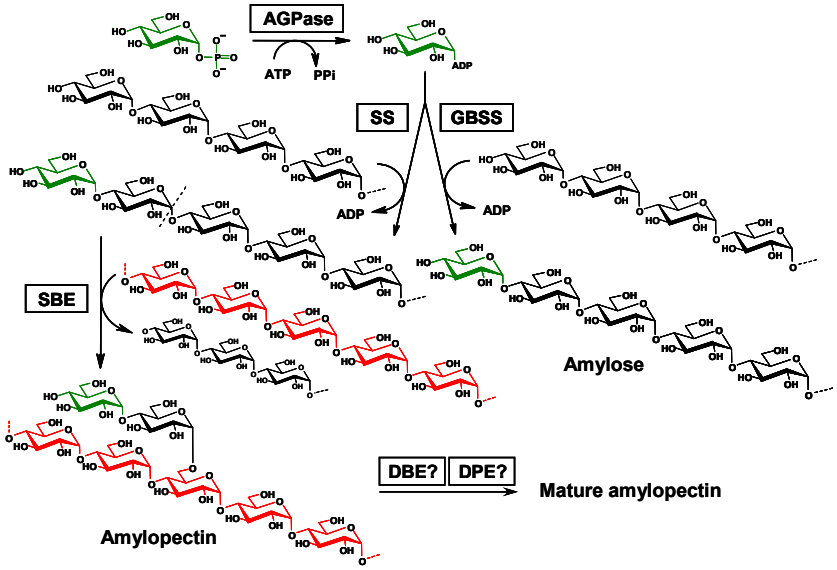


Figure 5. Overview of the first steps involved in starch biosynthesis.
The first committed step in starch biosynthesis is the formation of ADP-glucose in a reaction catalysed by ADPglucose pyrophosphorylase (AGPase). Subsequently soluble or granule bound starch synthases (SS and GBSS respectively) add the glucose moiety of ADPglucose to the non reducing end of an α-glucan polymer. In the final step α-1,6 branch points are introduced into the growing polymer by starch branching enzyme (SBE). Further processing of the amylopectin molecules possible involves debranching enzymes (DBE) and disproportionating enzyme (DPE).

In all these cases the defective gene product was shown to be the GBSSI protein. All the mutants had wild-type numbers of granules with normal organisation. It is therefore reasonable to assume that amylose synthesis occurs downstream of amylopectin. The GBSSI is found essentially completely within the granule matrix (Kuipers *et al.,* 1994; Tatge *et al.,* 1999). The second group of SS genes (designated SSI, SSII, SSIII and SSIV) is exclusively involved in amylopectin synthesis and their distribution within the plastid between the stroma and starch granules varies between species, tissue and developmental stage (Ball and Morell, 2003). Described so far SS can be divided into distinct classes on the basis of similarities in their amino acid sequences (Marshall *et al.,* 1996). They play distinct roles in the synthesis of amylopectin and exist either as soluble forms or as both soluble and minor granule-bound isoforms. All plants possess the granule-bound isoform GBSSI, whereas the SSI, SSII and SSIII isoforms are found in different combinations. Reduction in the activity of either the SSII or SSIII or a combination of the two in potato tubers lead to unique changes in amylopectin chain lengths (Edwards *et al.*, 1999). For example, recent studies with SSIII and SSIV in *Arabidopsis thaliana* suggest that SSIII can control the rate of starch synthesis in leaves (Zhang *et al.*, 2005) and that SSIV may be involved in a granule initiation pathway (Roldan *et al.*, 2007). The initial primers for the synthesis of glucan chains *in vivo* have not been identified. In the last step (3) rearrangements of the growing polysaccharide chain is catalysed by starch branching enzyme (SBE). Branch linkages within glucan chains of varying length are introduced by SBEs, which cleave internal α-(1/4) linkages and transfer the released reducing end to a C6 hydroxyl to create a new α-(1/6) linkage. In addition to SS and SBE, the debranching enzymes (DBE) also appear to be an essential component in the formation of semicrystalline amylopectin, because DBE mutants often lack or are reduced in amylopectin content. Instead large amounts of a soluble, glycogen-like glucan polymer accumulate (Zeeman *et al.,* 1998; Wattebled *et al.,* 2005). Studies of the *sta7* mutant in the green unicellular algae *Chlamydomonas reinhardtii*, which accumulates only phytoglycogen, lead to the proposal of the glucan-trimming model (Fig. 6) (Ball *et al.,* 1996; Mouille *et al.,* 1996). In this model elongation by SSs and branching by SBEs generate a highly branched precursor termed pre-amylopectin. The excess side chains are then removed by the action of DBE to generate the trimmed amylopectin which can form the semicrystalline structure of alternating lamellae. A revised trimming model was proposed based on data that showed that the disproportionating enzyme (DPE) also plays a role. The DPE acts by transferring (disproportionating) α-1,4 bonds from linear malto-oligosaccharides to an acceptor. The acceptor can be a polyglucan but can also be maltose or glucose. The *Chlamydomonas* mutant (*Sta11*) lacking the DPE accumulates less amylopectin and it contains shorter glucan chains than in wild type (Colleoni *et al.,* 1999a; Colleoni *et al.,* 1999b). The amount of linear malto-oligosacharides in the soluble fraction increased abnormally. Hence it was suggested that DPE transfers glucans among side chains in pre-amylopectin. Furthermore, malto-oligosaccharides trimmed of by DBE are transferred back to the maturing pre-amylopectin molecule (Myers *et al.,* 2000). Coordination of these activities is likely to be required to

produce the non-random clustered arrangement of glucan chains characteristic of amylopectin.

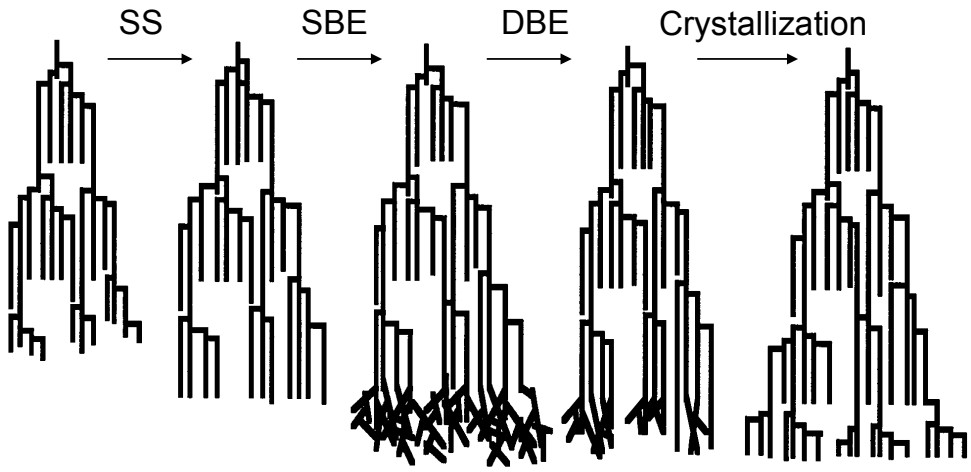


Figure 6. The glucan-trimming model.
The model is a proposal for explaining the interaction of enzymes involved in amylopectin synthesis. Existing side chains of amylopectin are elongated by Starch synthase (SS) followed by random branching by SBE. The highly branched glucans are then trimmed by DBE, and the semicrystalline structure of amylopectin is formed. Ball *et al*., (1996).

6.3 Degradation of starch

In photosynthetic tissue, starch degradation is a process that occurs to some extent all the time in balance with starch biosynthesis. In the amyloplast of storage tissues such as seeds and tubers, starch degrading enzymes are dormant until they are activated under germination or growth. Germination is accompanied by massive *de novo* synthesis of hydrolytic enzymes, so the situation is not comparable with the night-time degradation in leaves. The major end products of this starch mobilisation in the amyloplast are maltose, glucose and glucose-1-phosphate. The way in which the insoluble starch granule is degraded to release stored glucose is characterised in the endosperm of germinating cereals. In this tissue, the process of starch breakdown is initiated by the action of α-amylase (AMY), secreted into the non-living starchy endosperm (Fincher, 1989; Gubler *et al.,* 1999). AMY functions as an endoamylase and releases a mixture of branched and linear oligosaccharides that serve as substrates for other hydrolytic enzymes including debranching enzymes (which hydrolyze the α-1,6-bonds) and exoamylases (β-amylases). In *Arabidopsis* several different enzymes could potentially be involved in leaf starch degradation (Fig. 7).
Three proteins predicted to be AMY are encoded in the *Arabidopsis* genome. But T-DNA insertion mutants lacking AMY1, AMY2 or AMY3 have normal rates of starch degradation in leaves at night. Even a triple mutant lacking all three *α*-amylases also has normal rates of starch degradation (Yu *et al.*, 2005). Recent data from studies in *Arabidopsis* and potato are consistent with the idea that there might

be degradation of the granule surface by exo-amylases and DBEs. The loss of a specific isoform of the β-amylase (BAM3) in *Arabidopsis* (Kaplan and Guy, 2005) shows a SEX phenotype in their leaves. The same phenotype was observed in potato antisense plants with reduced expression of plastidial β-amylase (PCT-BMY1) in potato (Scheidig *et al.*, 2002). Similarly, the ABI4 transcription factor appears to be involved in the regulation of the transcripts of both GWD and the β-amylase BAM3 in *Arabidopsis* (Ramon *et al.,* 2007). Most higher plants contain four different DBEs: three isoforms of isoamylase and one limit–dextrinase, depending on their substrate specificity (Lloyd *et al.,* 2005). In *Arabidopsis*, the isoforms *At*ISA1 and *At*ISA2 form a multimeric complex with essential roles in starch biosynthesis, but not in degradation (Wattebled *et al.,* 2005). Knockout of the third isoamylase-type DBE *At*ISA3 lead to a strong SEX phenotype, indicating that this enzyme plays a major role in starch degradation (Delatte *et al.,* 2006).

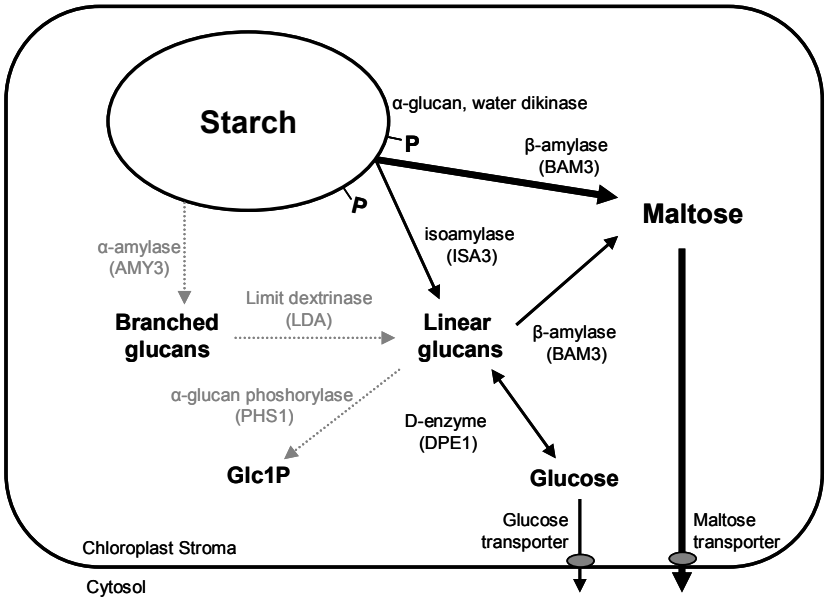


Figure 7. A model for the pathways of starch degradation in Arabidopsis chloroplast.
Phosphorylation of the granule surface by GWD may allow the direct action of BAM3 and ISA3. Linear glucan can be metabolized by DPE1, releasing glucose. Loss of any of these enzymes reduces starch breakdown and causes sex phenotype. AMY3 may also attack the granule to release branched glucans which can then be debranched in the stroma by LDA. In addition to further degradation by BAM, PHS1 may liberate glucose-1-phosphate to support chloroplast metabolism. The reactions catalyzed by AMY, LDA and PHS1 are shown in grey with broken lines, as loss of these proteins does not prevent normal rate of starch breakdown. The thickness of the arrows reflects an estimate of the relative fluxes. The importance of glucose transporter for starch breakdown has not yet been established (From Zeeman *et al.,* 2007).

The potato ISA3 seems to have a similar function as in *Arabidopsis* due to its preference for β-limit dextrins (small remaining parts of amylopectin after digestion with β-amylase). It was also shown, that recombinant potato ISA3 acts directly on the granular starch (Hussain *et al.,* 2003) suggesting that it might be involved in the initial release of glucans from the starch granule. Since down-regulation or

mutation of neither BAM nor ISA3 genes completely abolish starch breakdown, other enzymes are able to degrade starch in their absence (Kaplan and Guy, 2005).
Mutation in the debranching enzymes limit dextrinase (LDA) had no measurable effect (Delatte *et al.,* 2006). But when *isa3* and *lda* mutations were combined, the double mutant exhibited a more severe SEX phenotype than the *isa3* single mutant. These results indicate that in the absence of ISA3, LDA plays a role in starch degradation (Delatte *et al.,* 2006).
Linear oligosaccharides can be metabolised from their non-reducing ends by BAM to yield maltose and maltotriose. Maltose is believed to be the major form of carbon exported from the chloroplasts after remobilisation of starch by the chloroplast transporter MEX1 (Weise *et al.,* 2004; Niittyla *et al.,* 2004). Maltotriose can be converted via a glucantransferase reaction catalysed by DPE to maltopentaose and glucose. This enzyme is also involved in starch synthesis describe previously. In the *dpe1* mutant of *Arabidopsis*, maltotriose accumulates during the dark, and starch breakdown is reduced (Critchley *et al.*, 2001). Furthermore mutations or down-regulation of the cytosolic isoform, DPE2, results in high accumulation of maltose in leaves of *Arabidopsis* and potatoes and the plants show a SEX phenotype (Lloyd *et al.,* 2004; Chia *et al.,* 2004). Linear glucans can also potentially be metabolised by the chloroplast-localised α-glucan phosphorylase (PHS1). Phosphorylase liberates glucose-1-phosphate from the non-reducing end and the preferred substrate is oligosaccharides of five glucose residues or longer (Steup and Schächtele, 1981). Complete removal of the enzyme (PSH1) by insertional mutagenesis in *Arabidopsis* had only minor effect on leaf starch metabolism (Zeeman *et al.*, 2004).

6.4 Glucan, water dikinases (GWDs)

In addition to enzymes that cleave α-1,4 or α-1,6 linkages, also starch phosphorylating enzymes are typically required for normal starch mobilisation. Elucidation of the enzymatic mechanism responsible for starch phosphorylation came with the discovery of a granule-bound protein in potato termed R1 (Lorberth *et al.*, 1998). The protein was found to be a glucan, water dikinase termed GWD which transfers the β-phosphate group of ATP to a small proportion of the glucose residues of amylopectin chains (Ritte *et al.,* 2002; Mikkelsen *et al.,* 2004; Ritte *et al.,* 2006). Antisense repression of GWD resulted in an up to 90% reduction of the phosphate groups covalently linked to the amylopectin component of both leaves and tubers. It also caused starch to accumulate in leaves (Fig. 8) and prevented the cold-induced sweetening of stored tubers. Subsequent work revealed an *Arabidopsis* mutant selected because of the accumulation of very high levels of starch in its leaves (the *sex1* mutant) carrying a mutation in the gene encoding GWD. The *sex1* mutant alleles showed significant homology with the starch granule-bound enzyme R1 isolated from potato and the mutant also lacks phosphate groups in the amylopectin component of starch (Yu *et al.,* 2001). Analysis of different alleles of the *sex1* mutant showed that the starch content was 3-7 times higher than in the wild type and that there was a good inverse correlation between the degree of phosphorylation and the severity of the SEX phenotype (Yu *et al.,* 2001). *sex1* mutants also accumulate elevated levels of starch in seeds, flowers and root tips (Caspar *et al.*, 1991) and have reduced growth rates, due to the inability to mobilize

starch during the night. These results suggest that GWD mediated phosphorylation of amylopectin is required for starch breakdown in both photosynthetic and non-photosynthetic tissues.

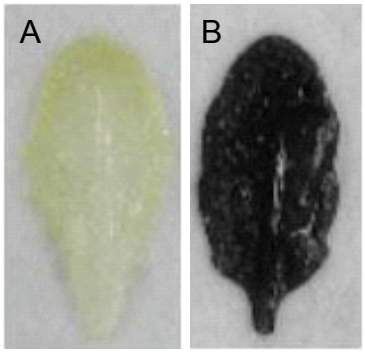


Figure 8. The *GWD*-knockout phenotype:
A. Normal starch degradation during dark period in wild-type *Arabidopsis* leaves, monitored by iodine staining of a leaf harvested after a 12 h dark period.
B. Leaf from a *sex1* genotype, in which *GWD* has been knocked out, iodine staining results in dark blue colouring of a leaf harvested after a 12 h dark period, indicating incomplete starch degradation (Blennow *et al.,* 2002).

6.4.1 Catalytic mechanism.

Sequence analyses of GWD have revealed that the C-terminus of GWD shows identity with the nucleotide-binding domain and phosphohistidine domain of bacterial PPS (pyruvate, water dikinase) and PPDK (pyruvate, phosphate dikinase) which transfers phosphate from ATP to pyruvate and water in a dikinase-type reaction. The enzymatic function of the GWD protein from potato was determined by heterologous expression in *E. coli* and it was shown that GWD is a dikinase which can transfer phosphate from ATP to starch and water (Ritte *et al.,* 2002; Mikkelsen *et al.,* 2004).
GWD phosphorylates α-glucans by a dikinase-type reaction mechanism, which can be divided into two steps (Fig. 9).

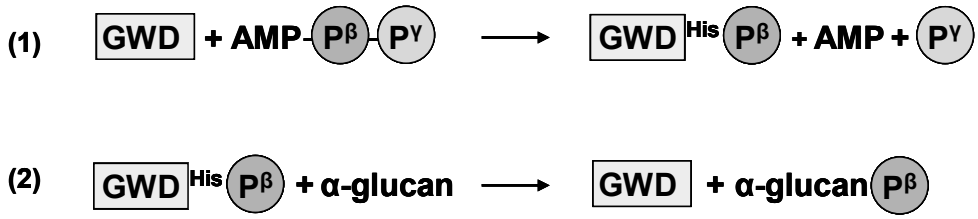


Figure 9: Phoshorylation mechanism of GWD.
During catalysis, the β-phosphate of ATP is transferred to a histidine residue of GWD generating a stable phosphohistidine intermediate. Subsequently this phosphate is transferred to the glucan molecule.

The first step involves autophosphorylation of GWD and a stable phosphohistidine intermediate is formed between the β-phosphate from ATP and GWD (1). In the next step the β-phosphate is transferred to a glucosyl residue in the α-glucan (2). In the first step GWD binds ATP and then γ-phosphate is transferred to water and the β-phosphate is transferred to a phosphohistidine intermediate. The His-992 of potato

GWD undergoes autophosphorylation during catalysis and replacing this histidine residue with an alanine results in completely inactive enzyme (Mikkelsen *et al.,* 2004). The first reaction also requires the ATP-binding domain which is located in the C-terminal end (Fig. 9) (Mikkelsen *et al.,* 2005). In the second step, the histidine bound phosphate is transferred to a glucosyl residue in the α-glucan. Circular dichroism experiments have shown that during autophosphorylation of potato GWD1, the enzyme undergoes major conformational changes. A hypothetical model has been suggested illustrating a domain rearrangement during the phosphorylation process (Mikkelsen and Blennow, 2005). After autophosphorylation of the catalytic histidine domain rearrangements occurs and the tandem SBDs with bound glucan substrates are in close proximity to the phosphorylhistidine and thereby initiate the phoshorylation of the glucan substrate (Mikkelsen and Blennow, 2005).

6.4.2 Homologues

In the *Arabidopsis* genome three isoforms of GWD have been identified and all homologues contain a conserved nucleotide binding domain and a phosphohistidine domain (Mikkelsen *et al.,* 2004; Baunsgaard *et al.,* 2005; Kötting *et al.,* 2005)(Fig. 10). The potato GWD shows 66.3% amino acid sequence identity with the *At*GWD1. *At*GWD2 has 50% amino acid sequence identity to *At*GWD1, compared to the *At*GWD3 which has a rather weak homology to the two other *Arabidopsis* homologues of about 30% (Yu *et al.,* 2001). The sizes differ between the homologues as well as the domain organisation (Fig. 10). A predicted chloroplast transit peptide is found at the N-terminus in all homologues except the *At*GWD2 (Yu *et al*., 2001) and suggestions have been made that it might be a cytosolic isoform to *AtGWD1*. It appears that *At*GWD2 is expressed in the companion cells of the phloem and maybe involved in degradation of sieve-element starch or soluble glucans in the phloem (Glaring *et al.*, 2007). Another interesting variation is the starch binding domains located in the N-terminus. In potato GWD, *At*GWD1 and *At*GWD2 the starch binding domain belong to family CBM45, whereas in the *At*GWD3 it belongs to the well characterised CBM20 family. A detailed description of the CBM classification system and the function of the starch-binding domains can be found in Paper IV.

The presence of GWD appears to be a ubiquitous feature of all organisms accumulating semicrystalline storage polysaccharides. Homologues of GWD can be found in many plants (e.g *Arabidopsis*, barley, maize, wheat, rice, potato, rape, tomato and tangerine) (Ritte *et al.,* 2000a; Yu *et al.,* 2001) and in the green algae *Chlamydomonas reinhardtii* and *Ostreococcus tauri*, the red algae *Cyanidiumschyzon merolae* and the apicomplexan parasites *Toxoplasma gondii* and *Cryptosporidium parvum* (Mikkelsen *et al.,* 2004; Ral *et al.,* 2004; Coppin *et al.,* 2005).

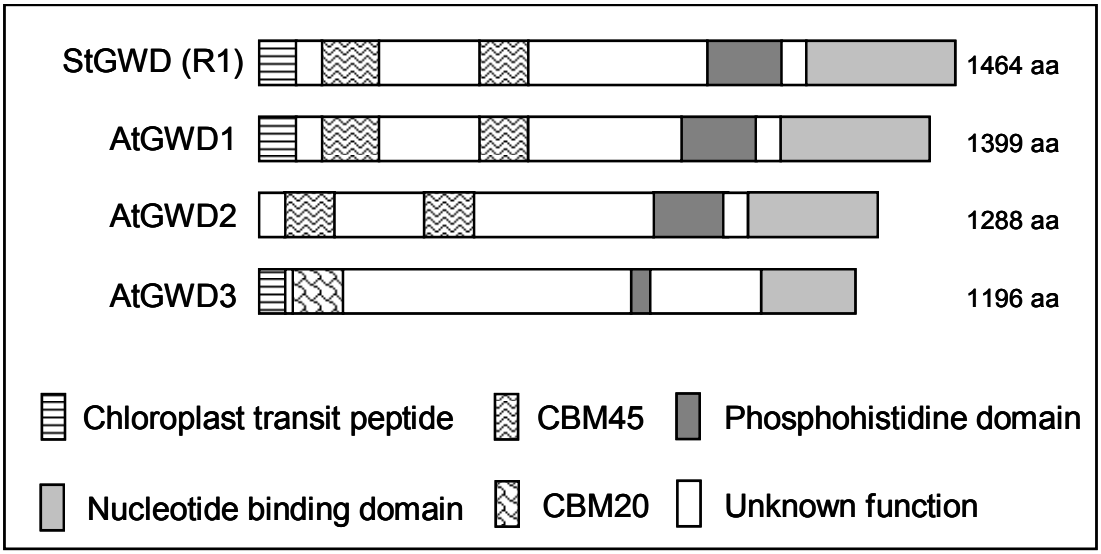


Figure 10: Domain organisation in plant dikinases.
Chloroplast transit peptides are found in potato GWD, *At*GWD1 and *At*GWD3, whereas the localisation of *At*GWD2 is extracellular. Three of the GWD homologues posses a tandem repeat of the CBM45 and one homolog, *At*GWD3, posses a CBM20. Sequence analyses have revealed that the C-terminus shows homolog to a nucleotide binding domain and a phosphohistidine domain. The sizes of the full length enzymes in amino acid residues (aa) are indicated.

6.4.3 Substrate specificity

The *in vitro* activity of GWD depends on the glucan substrate. Glycogen and amylose are both very poor substrates (Ritte *et al.,* 2002; Mikkelsen *et al.,* 2004). Longer amylopectin chains are more phosphorylated than shorter chains (Blennow *et al.,* 1998) and analysis of phosphorylated substrates demonstrated that phosphate was incorporated in chains with DPs between 30 and 100 (Mikkelsen *et al.,* 2004). The preference for longer chains offers an explanation for why the cereal starches, which have shorter amylopectin chain lengths, are only slightly phosphorylated, despite the presence of comparable levels of GWD enzyme (Ritte *et al.,* 2000a). The *in vitro* activity is strongly dependent on the amylopectin chain length and a twenty-fold increase was observed when the mean DP was increased from 27.9 to 29.5 (Mikkelsen *et al.,* 2004). GWD displays significant activity on starch particles compared with solubilised starch (Ritte *et al.,* 2004; Kötting *et al.,* 2005). *At*GWD1 phosphorylates α-glucans preferentially in the C6 position and *At*GWD3 phosphorylates exclusively the C3 position (Baunsgaard *et al.,* 2005; Kötting *et al.,* 2005; Ritte *et al.,* 2006). Also difference in substrate specificity has been found between the *At*GWD1 and *At*GWD3. *In vitro* results show that *At*GWD3 is only able to phosphorylate pre-phosphorylated glucans (Baunsgaard *et al.,* 2005) and therefore *At*GWD3 is also named PWD (phosphoglucan, water dikinase). Plants lacking *At*GWD1 contain no detectable phosphate. This implies that *At*GWD3 acts downstream of *At*GWD1 and depend on the prior action of *At*GWD1 (Ritte *et al.,* 2006). Actually,

GWD1 can only phosphorylate C6 *in vivo* even though it phosphorylates at both positions *in vitro* (Ritte *et al.,* 2006).

6.4.4 Regulation of activity

Phosphorylation of potato amylopectin occurs as starch is synthesised (Nielsen *et al.,* 1994), resulting in a low level of phosphorylation throughout the granule. However, GWD is reported to be more active during periods of starch breakdown. GWD has been recovered in the fraction of starch-associated proteins from granules isolated from starch degrading (darkened) leaves, whereas the proportion of granule-bound GWD was negligible during starch synthesis in the light (Ritte *et al.,* 2000b; Reimann *et al.,* 2002).

In vitro phosphorylation assays using recombinant potato GWD revealed that granules from darkened leaves were phosphorylated at a much higher rate than granules from illuminated leaves (Ritte *et al.,* 2004).

Storage starch in turions (dormant vegetative buds) of the duckweed *Spirodela polyrhiza* is degraded in response to red light. Irradiation of dark-adapted turions with continuous red light resulted in *in vivo* starch degradation. Analysis of GWD partitioning revealed that after illumination the level of granule-bound GWD decreased. This dissociation preceded the decrease in starch content (Reimann *et al.,* 2002). Furthermore it was shown (Reimann *et al.,* 2004) that irradiation of dark-adapted turions results in autophosphorylation of GWD bound to the starch granules. The irradiation of the turions also resulted in increased phosphorylation of isolated starch granules upon incubation with ATP (Reimann *et al.,* 2004; Reimann *et al.,* 2007). They suggested that starch phosphorylation resulting from irradiation enhances the binding of starch-degrading enzymes such as α-amylase to the granule surface and thereby initiates the degradation of the starch.

The unicellular green alga *Chlamydomonas reinhardtii* has been used a model system for incorporation of ^{32}P and detection of transient starch phosphorylation. In illuminated cells, phosphate is incorporated into starch at a constant rate without significant turnover. In contrast, phosphorylation in darkened cells exceeded that in illuminated cells within the first 30 minutes and was accompanied by significant turnover (Ritte *et al.,* 2004). This suggests that starch phosphorylation during degradation is transient and that the activity of GWD is increased as a response to darkness in *Chlamydomonas reinhardtii*. Furthermore, in *Chlamydomonas reinhardtii* the level of phosphate incorporated into starch during the dark started dropping after only 90 minutes in the dark (Ritte *et al*., 2004), perhaps suggesting that GWD activity and/or binding capacity is not constant during the dark period. To further examine transient starch phosphorylation the outer surface of potato leaf starch granules was analysed after a short incubation with isoamylase, leading to the release of glucan chains exposed at the granule surface (Ritte *et al*., 2004). Analysis of the released glucan chains showed that both single and double phosphorylated chains were highly enriched in starches from darkened leaves compared to illuminated leaves. This enrichment was only poorly reflected in the total phosphate content of the granules, suggesting that the overall content of starch-bound phosphate reflects phosphate incorporated during biosynthesis.

Recently evidence was presented that the potato GWD is strongly regulated via redox activation by an internal disulfide bond between the two cysteines C1004 and C1008 was reported (Mikkelsen *et al.,* 2005). Treatment of GWD with the oxidizing agent $CuCl_2$ resulted in complete inactivation of the enzyme. The activity could subsequently be restored by incubation with the reducing agent DTT and purified thioredoxin. The enzyme exists in its inactive oxidized form when attached to the starch granule in the dark, whereas the soluble form is reduced and fully active. A mutant GWD enzyme (C1008S) incapable of forming the disulfide bridge was permanently active over a wide redox range. Analysis of granule-bound and soluble GWD from dark-adapted and illuminated potato plants showed that GWD was primarily bound to the granules in the inactive, oxidized form during darkness and that soluble GWD was reduced and fully active (Mikkelsen *et al.,* 2005). The CFATC motif found in potato GWD was not conserved in all GWD homologues, indicating that they can be regulated by other mechanisms or not redox regulated (Mikkelsen *et al.,* 2005). Further work is required to resolve the impact on GWDs reversible binding properties during phosphorylation. Several enzymes might act in synergy and there are several ways all these enzymes can be regulated in a dynamic organelle like the plastids. GWD could also be regulated by other mechanisms besides redox level *in vivo.*

6.5 Carbohydrate binding modules and classification

Glucosidic bonds from different polysaccharides such as cellulose, xylan and starch are hydrolysed by cellulases, xylanases and amylases, respectively. Glycoside hydrolases attack polysaccharides relatively inefficiently as their targets are often inaccessible to the active site of the attaching enzymes. Many microbial enzymes that catalyze plant cell wall hydrolysis have a modular structure in which a noncatalytic carbohydrate binding module (CBM) targets the enzyme to specific polysaccharides. A CBM is defined as a contiguous amino acid sequence within a carbohydrate-active enzyme with a discrete fold having carbohydrate-binding activity. The term CBM is used to describe all of the non-catalytic carbohydrate binding modules derived from glycoside hydrolases and they are classified into different families according to the CAZy classification system (see http://www.cazy.org/fam/acc_CBM.html). The CAZy database describes the families of structurally related catalytic and carbohydrate-binding modules (or functional domains) of enzymes that degrade, modify, or create glycosidic bonds. Many CBMs have been identified and are currently classified into 53 families on the basis of amino acid sequence similarity. CBMs with starch-binding activity are also called starch-binding domains (SBDs) and they are found in nine of the CBM families (20, 21, 25, 26, 34, 41, 45, 48 and 53), sharing a very similar fold: a β-sandwich fold with an immunoglobulin-like topology.

6.5.1 Family CBM20

There are currently 276 entries in CBM20 in CAZy. The CBM20 modules occur together with a variety of bacterial and fungal catalytic domains and most frequently they are found in the three glycoside hydrolase (GH) families; GH13: the α-amylase family, GH14: the β-amylases and GH15: glucoamylases (Henrissat and

Davies, 1997). A few CBM20s are found in plants and include the GWD3 (Baunsgaard *et al.,* 2005) and the 4-α-glucanotransferase from GH77 (Steichen *et al.,* 2008) and the human CBM20 Laforin (Minassian *et al.,* 2000) and genethonin-1 (Janecek, 2002) are also non-hydrolytic enzymes.

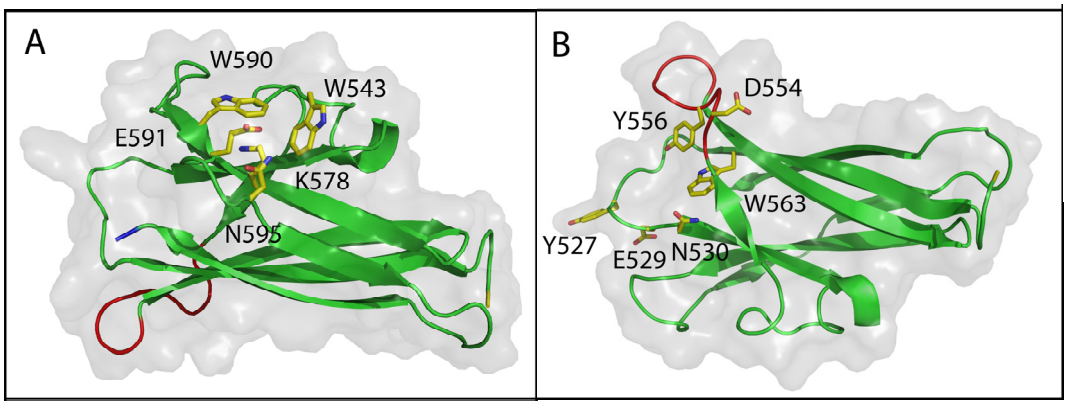


Figure 11: A cartoon representation of the *Aspergillus niger* glucoamylase CBM20 structure (PDB code 1kul)
A: Binding site 1. **B**: binding site 2 The cartoon is coated by a transparent molecular surface representation to give a topological perspective of the binding sites and the N- and the C-termini are coloured in yellow and blue, respectively. The view in panel B is roughly rotated 180° along the long axis of the molecule and the flexible loop showing the largest conformational change upon ligand binding is coloured in red. The aromatic ligand binding residues and selected residues implicated in ligand interaction at each site are shown as sticks.

CBM20 modules are between 90-130 residues long and detailed analysis of amino acids sequences revealed that there are no invariant residues in the CBM family 20 (Machovic *et al.,* 2005; Machovic and Janecek, 2006). Sequence alignment together with three-dimensional structures provided support for two independent glucan binding sites on the CBM20 scaffold. The three-dimensional structures of the *A. niger* glucoamylase SBD has been solved both in the free state (Sorimachi *et al.,* 1996) and bound to β-CD (Sorimachi *et al.,* 1997). The SBD forms an open-sided, distorted, β-barrel structure with in total eight β-strands. Binding site 1 contains two conserved tryptophan residues, which are W543 and W590 in the *A. niger* GA-SBD (Williamson *et al.,* 1997). Site two includes the two aromatic residues Y527 and Y556, but is in general less conserved and some CBM20 no binding has been observed in the binding site 2. Binding site 1 is small and forms a ridged hydrophobic binding platform believed to function in initial recognition of the substrates and is found facing away from the catalytic part of the enzymes. Binding site 2 is more flexible and undergoes conformation change upon binding and this flexibility could be used to guide single amylose chains closer to the catalytic site. Furthermore detailed description of the CBM20 is given in the results and in Papers I, III and IV.

7 Experimental work

7.1 Constructs, expression and purification

7.1.1 GWD3-SBD *E. coli* Gateway

Plasmid construction

A pYES2.1/V5-His-TOPO plasmid (Invitrogen, Paisley, UK) containing the full-length GWD3 cDNA (GenBank accession no. AY747068) was used as a PCR template. Phusion™ High-Fidelity DNA Polymerase (New England Biolabs, Hitchin, UK) was used to amplify Asp^{68}-Thr^{170} (Fig. 7.1) with primer P1 and P2 (Appendix table 9.1).

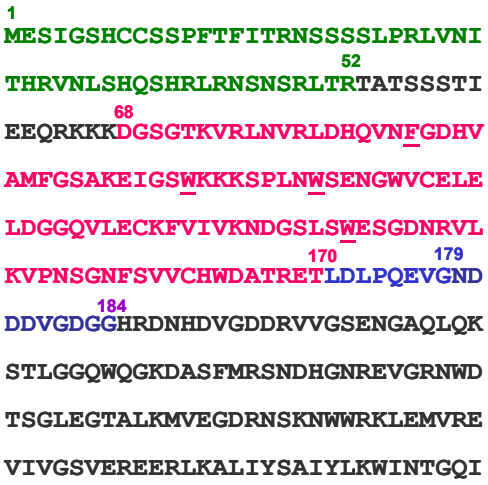


Figure 7.1 *Arabidopsis thaliana* GWD3 amino acid sequence (1-300).
The chloroplastic transit peptide sequence (1-52) is shown in green. Previously defined domain annotation of GWD3-SBD (68-170) is shown in red. The additional amino acid stretch used to defined domain annotations with a longer C-terminal region (171-184) is shown in blue. Mutations made in the pDONR201 are indicated with underlined letters.

The *attB* PCR product was subsequently by recombination cloned into pDONR201 vector by recombination using the Gateway BP reaction according to manufacturer's instructions (Invitrogen). The insert from the Gateway entry clone was transferred by recombination into pDEST14 (without tag), pDEST15 (with an N-terminal Glutathione S-transferase **(**GST) tag**)** and pDEST17 (with an N-terminal His-tag**)** (Table 5.1) to make the Gateway expression plasmids..

Table 5.1: Vectors used in this study

Expressi on host	Vector	Fusion Peptide	Fusion Tag
E. coli	pDEST14	-----	------
E. coli	pDEST15	N-terminal	GST
E. coli	pDEST17	N-terminal	6xHis
E. coli	pET41	C-terminal N-terminal	6xHis and GST
E. coli	pET28*	C-terminal	6xHis
P. pastoris	pPicZα B	C-terminal	6xHis and c-myc epitope (EQKLISEEDL)

Vectors are from Invitrogen except *which is from Novagen

The plasmids were electroporated into *E. coli* BL21A1 (Invitrogen) for expression and into *E. coli* TOP10 (Invitrogen) for long time storage and correct inserted DNA was verified by sequencing

Expression and purification

E. coli strain BL21A1 harbouring pDEST14 with GWD3-SBD were grown at 37 ºC in 1 L LB medium supplemented with 100 µg/mL ampicillin to OD600~ 0.4. L-arabinose was added to a final concentration of 0.2% and the cultivation was continued at 28 ºC or 37 ºC. Samples were collected every hour to verify expression by SDS-PAGE. No clear overexpression was seen on the SDS-PAGE and since no antibodies were available for verifying soluble proteins expression, another approach was chosen.

E. coli strain BL21A1 harbouring pDEST15 with GWD3-SBD were grown under the same conditions as above, but antibodies against the GST tag were used to confirm GWD3-GST fusion expression. Cells grown for 5 hours at 28 ºC were pelleted by centrifugation at 5000 x g for 20 min at 4 ºC. Larger scale expression was also done in a 5 L Bioreactor (BIOSTAT B), to evaluate if the controlled conditions would increase yield to enable unambiguous detection. After harvest the cells were resuspended in 10 mL (500 mM Tris-HCl, 30 mM EDTA, 0.5 mM phenylmethylsulphonyl fluoride (PMSF), 5 mM dithiothreitol (DTT) pH 7) and disrupted by sonication. The extracts were clarified by centrifugation 12000 x g for 30 min at 4 ºC.

The samples were loaded with a flow of 0.2 mL/min on a 1 mL GSTrap-column (GE-Healthcare, Uppsala, Sweden) and washed with 20 CV of washing buffer (140 mM NaCl, 2,7 mM KCl, 10 mM Na_2HPO_4, 1.8 KH_2PO_4, 5 mM DTT pH 7.3) with a flow of 0.7 mL/min and eluted with a gradient of 10 CV in elution buffer (50 mM Tris, 20 mM GSH, 200 mM NaCl and 10% glycerol, pH 8.0) and a flow of 0.5 mL/min. Eluted fractions were analysed by SDS-PAGE. Eluted protein fractions from the GSTrap-column were dialysed against 10 mM Tris pH 8.0 and loaded on a 1 mL fast flow HiTrap Q strong anion exchanger (GE Healthcare) equilibrated with 20 mM Tris pH 8.0 and a flow of 0.5 mL/min. Protein were eluted with a 10 CV gradient of 20 mM Tris, 1 M NaCl pH 8.0 and a flow of 0.5 mL/min. Eluted fractions were analysed by SDS-PAGE. Another batch of expressed protein was after a GSTrap purification dialysed against 10 mM HEPES pH 7.5 and loaded on an XK 16/100 column (GE Healthcare) packed with P30 acrylamide gel filtration resin (BioRad, Hercules, Ca) and equilibrated with 10 mM HEPES, 250 mM NaCl pH 7.5 with a flow of 0.15 mL/min. Eluted fractions were analysed by SDS-PAGE and another purification step was needed. Dialysed protein were added 5% (w/v) $(NH_4)_2SO_4$ and mixed with 2 mL β-CD linked Sepharose resin two hours at 4 ºC. The resin was pelleted by centrifugation 5 min at 1000 x g and 4 ºC and washed with 20 mM HEPES, 250 mM NaCl, 5% (w/v) $(NH_4)_2SO_4$ pH 7.4 twice. To elute bound protein 20 mM HEPES, 10 mM β-CD pH 7.4, was added to the pellet resin and samples were analysed by SDS-PAGE.

E. coli strain BL21A1 harbouring pDEST17 with GWD3-SBD were grown at 37 ºC in 1 L LB medium supplemented with 100 µg/mL ampicillin to OD_{600} ~0.4. Protein expression was induced with 0.2% L-arabinose and the culture was grown at 28 ºC

for 5 hours and 200 rpm. Cells were harvested by centrifugation, 5000 x g for 20 min at 4 ºC and cells were resuspended in (20 mM HEPES, 20 mM imidazole, 500 mM NaCl, 5% glycerol, 0.5 mM PMSF and 5 mM DTT pH 7.4). Larger scale expression was also done in a 3 L Bioreactor. The resuspended cells were sonicated, followed by a 12000 x g centrifugation for 30 min at 4 ºC.

Sample was loaded with a flow of 0.5 mL/min on a 1 mL HisTrap column (GE Healthcare). The column was washed with 20 CV of washing buffer (20 mM HEPES, 20 mM imidazole, 500 mM NaCl, 5% glycerol pH 7.4) and a flow of 1 mL/min and eluted with a flow of 0.5 mL/min in elution buffer (20 mM HEPES, 500 mM imidazole, 500 mM NaCl, 5% glycerol) and a gradient of 20 CV.

Unbound protein was collected and applied with a flow of 0.5 mL/min on a 6 mL β-CD linked Sepharose column equilibrated with 50 mM sodium acetate, 250 mM NaCl, 5% (w/v) $(NH_4)_2SO_4$ pH 5.5 and protein was loaded with a flow of 0.3 mL/min and recirculated three times. The column was washed until the A_{280} signal reached the baseline with a flow of 1 mL/min and protein was eluted with 50 mM sodium acetate, 10 mM β-CD pH 5.5 and a flow of 0.5 mL/min. No change in absorbance occurred during elution and the flow through from the β-CD linked Sepharose column was dialysed against 20 mM HEPES pH 7.4 and added 10% (w/v) $(NH_4)_2SO_4$ and reloaded on a 6 mL β-CD linked Sepharose column equilibrated with 20 mM HEPES, 250 mM NaCl, 5% (w/v) $(NH_4)_2SO_4$ pH 7.4 and same purification methods were applied as before. Elution with 20 mM HEPES, 10 mM β-CD pH 7.4 did not result in any absorbance change either. The flow through from the β-CD linked Sepharose column was applied to a HiLoad superdex G75 (16/60) gel filtration column (GE healthcare) equilibrated with 10mM HEPES, 250 mM NaCl, 5% glycerol pH 7.4. The sample was loaded with a flow of 0.5 mL/min. Fractions were analysed by SDS-PAGE.

Raw extract from another expression was sonicated and centrifuged as described above but the sample was loaded with a flow of 0.5 mL/min on a 1 mL NiNTA column (Qiagen, Hilden, Germany). The column was washed with 20 CV of washing buffer (20 mM HEPES, 20 mM imidazole, 500 mM NaCl, 5% glycerol pH 7.4) and a flow of 1 mL/min and eluted with a flow of 0.5 mL/min in elution buffer (20 mM HEPES, 500 mM imidazole, 500 mM NaCl, 5% glycerol) and a gradient of 20 CV. The flow through from the NiNTA column was dialysed against a 10 mM sodium acetate buffer pH 5.5 and loaded on a 6 mL Resource S column, cation exchange column (GE-Healthcare) equilibrated in the same buffer with a flow of 0.5 mL/min. The column was washed with 10 CV and a flow of 2 mL/min and eluted with a gradient of 10 CV and a flow of 0.5 mL/min with 10 mM sodium acetate, 500 mM NaCl buffer pH 5.5. The protein was collected in the flow through as no protein was bound to the column. The flow through from the Resource S column was dialysed against a 10 mM BICINE buffer pH 8.5 and loaded on a 6 mL Resource Q anion exchange column (GE-Healthcare) equilibrated in the same buffer with a flow of 0.5 mL/min. The column was washed with 10 CV and a flow of 2mL/min and eluted with a gradient of 18 CV and a flow of 0.5 mL/min with 10 mM BICINE, 500 mM NaCl buffer pH 8.5. Fractions of bound protein was analysed on SDS-PAGE. Bands of expected molecular weight were excised from the SDS-PAGE gel, and trypsin digested. Matrix assisted laser desorption

ionisation-time of flight (MALDI-TOF) mass spectrometry analysis confirmed that the band originated from the correct recombinant protein, but also showed that the produced polypeptide lacked the C-terminal His-tag.

7.1.2 GWD3-SBD Site-directed mutagenesis

Four single mutants (F86A, W102T, W110T and W142T), two double mutants (F86A/W142T and W102T/ W142T) and one triple mutant (F86A/W102T/W142T) were prepared using QuickChange XL site-directed mutagenesis kit (Stratagene) with primers listed (Appendix table 9.1). The mutations were carried out according to the recommendations of the manufacturer, and this was followed by recombination in the pDEST15 expression vector.

7.1.3 GWD3-SBD *E. coli* pET28a(+)

Plasmid construction

A pYES2.1/V5-His-TOPO plasmid containing the full-length GWD3 cDNA (GenBank accession no. At5g26570) was used as a PCR template to amplify amino acid sequence (68-179) with primer P11 and P12 (Appendix table 9.1) and amino acid sequence (68-184) with primer P11 and P13 (Appendix table 9.1). The PCR products were purified using QIAquick Gel extraction kit (QIAgen), digested with *nco*I and *xho*I (New England Biolabs) and subsequently ligated into pET28a(+) vector (Novagen, Madison, WI). The plasmids were transformed into *E.coli* BL21A1 (Invitrogen) and correct inserted DNA was verified by sequencing.

Expression and purification

The *E. coli* strain BL21(DE3) harbouring pET28a(+) with the GWD3-SBD (68-179) and GWD3-SBD (68-184) were grown at 30 ºC in LB medium with 50 µg/mL kanamycin to OD_{600} ~ 0.5 at 600 nm. Isopropyl β-D-thiogalactopyranoside was added to a final concentration of 250 µM and the culture was incubated at 17 ºC for 20 hours. The cells were pelleted by centrifugation at 12,000 *g* for 20 min at 4 ºC, resuspended in 20 mL buffer A (20 mM sodium phosphate, 0.5 M NaCl, 10 mM imidazole, pH 8) and disrupted by sonication. The extracts were clarified by centrifugation at 12000 x g for 30 min at 4 ºC.

Sample was loaded on a 1 mL Ni-NTA-column (QIAgen), washed with 15 column volumes and eluted with 60% buffer B (20 mM sodium phosphate, 0.5 M NaCl, 500 mM imidazole, and 10% glycerol, pH 8). After this first purification step both GWD3-SBDs were ~50-70% pure, therefore the eluted fraction was directly applied to a 6 mL β-CD linked Sepharose column for a second purification step. Bound protein was washed with 2 CV buffer (50 mM NaCl, 250 mM NaCl pH 7.5) and eluted in 50 mM HEPES, 10 mM β-CD, pH 7.5. Fractions were analysed by SDS-PAGE and combined for buffer exchange by dialysis against 50 mM HEPES, pH 7.5 (molecular weight cut-off, MWCO 3.5, Spectra/Por®) and stored at 4 ºC. Protein concentration in the enzyme preparation was measured spectrophotometrically at 280 nm using the theoretically calculated molar extinction coefficient (ε =28810 M^{-1} cm^{-1}). SDS-PAGE analysis was carried out to assess purity.

7.1.4 GWD3-SBD *P. pastoris* pPICZα B

Plasmid construction

The plasmid encoding GWD3-SBD in *P. pastoris* was constructed by PCR amplification using primer P14 and P15 (Appendix table 9.1). The PCR product was digested with *Sfi*I and *Xba*I and ligated into pPICZα B (Invitrogen) digested with the same enzymes. The plasmids were transformed into DH5α by electroporation. Positive transformants were selected on low salt LB plates with 25 µg/mL zeocin and propagated. Plasmids were purified using miniprep, spin columns (QIAgene) and sequenced. After sequencing the plasmids with correct inserted DNA were linearized by *Pme*I and transformed into *P. pastoris* strain X33 by electroporation. Transformants were selected on YPDS plates (1% (w/v) yeast extract, 2% (w/v) peptone, 2% (w/v) glucose, 1 M sorbitol, 2% (w/v) agar and 100 µg/mL zeocin). Transformants secreting GWD3-SBD were screened on MM plates (1.34% (w/v) yeast nitrogen base, 4 x 10^{-5}% (w/v) biotin, 0.5% (v/v) methanol and transferred onto a nitrocellulose membrane using replica plating. Positive transformants were detected by Western blotting using the c-myc antibody.

Expression and purification

Small scale pilot expression tests were made to optimise expression. Different temperatures during induction, addition of casamino acids and different shake flasks options were evaluated and to determine the optimal harvest time to attempt to minimise proteolytic cleavage of recombinant proteins, and samples were collected from different time points during induction and analysed by SDS-PAGE or Western blot. Optimal expression conditions were as follows: transformants were grown in 0.5 L BMGY (1% (w/v) yeast extract, 2% (w/v) peptone, 100 mM potassium phosphate, pH 6.0, 1.34% (w/v) yeast nitrogen base, 4 x 10^{-5}% (w/v) biotin, 1% (v/v) glycerol) at 30˚C for 18 hours in 3 L Erlenmeyer baffled flasks to OD_{600} ~ 2. Cells were harvested (20˚C, 1500 rpm, 8 min) and resuspended in 1L BMMY (1% (w/v) yeast extract, 2% (w/v) peptone, 100 mM potassium phosphate, pH 6.0, 1.34% (w/v) YNB, 4 x 10^{-5}% (w/v) biotin and 0.5% (v/v) methanol) to induce expression. The culture was incubated for 18 hours at 20˚C shaken with 200 rpm. Cells were harvested by centrifugation (4˚C, 9000 rpm, 20 min) and the supernatant were collected.

2 tablets of protease inhibitor cocktail (Roche, Mannheim, Germany) were added and the pH of the supernatant were adjusted to 7.4 by sodium hydroxide and 10 mM imidazole, 500 mM NaCl, and after stirring the supernatant was filtered through 0.22 µm filter. The supernatant was loaded on a 1 mL HisTrap column (QIAgen) and a flow of 1 mL/min, washed with 15 column volumes and eluted with a gradient of 20 CV from 0 - 100% elution buffer (20 mM sodium phosphate, 0.5 M NaCl, 500 mM imidazole, and 10% glycerol, pH 8). Since no fraction contained pure protein, the fractions were combined and dialysed against 10 mM sodium acetate pH 5.5 for a second purification step. Combined samples were applied with a flow of 0.5 mL/min on a 6 mL β-CD linked Sepharose column equilibrated with 50 mM sodium acetate, 250 mM NaCl, 5% (w/v) $(NH_4)_2SO_4$ pH 5.5 and protein was loaded with a flow of 0.3 mL/min. The column was washed until the A_{280} signal reached the baseline with a flow of 1 mL/min and protein was eluted with 50

mM sodium acetate, 10 mM β-CD pH 5.5 and a flow of 0.5 mL/min. No change in absorbance occurred during elution and the flow through from the β-CD linked Sepharose column was dialysed against 20 mM HEPES pH 7.4 and added 10% (w/v) $(NH_4)_2SO_4$ and reloaded on a 6 mL β-CD linked Sepharose column equilibrated with 20 mM HEPES, 250 mM NaCl, 5% (w/v) $(NH_4)_2SO_4$ pH 7.4 and same purification methods was applied as before in order to anticipate a pH dependent affinity. Elution with 20 mM HEPES, 10 mM β-CD pH 7.4 did not yield any absorbance change either and the flow through from the β-CD column were dialysed against a 10 mM HEPES pH 7.5 buffer. The dialysed flow through was loaded with a flow of 0.7 mL/min on a Superdex G75 HiLoad 16/60 gel filtration column (GE Healthcare) equilibrated with 10 mM HEPES, 200 mM NaCl pH 7.5 and eluted peaks were analysed for GWD3-SBD by SDS-PAGE. Although no calibration of the column was done, it could be concluded from the column guidelines that the elution time of the peak containing GWD3-SBD corresponded to about 3 times higher molecular weight than the theoretically expected (13.9 kDa), and a reducing treatment of the GWD3-SBD with DTT was performed. Samples were dialysed against 10 mM HEPES pH 7.5 and concentrated by ultrafiltration spin columns and DTT was added to a final concentration of 10 mM and the sample was left for 3 hours at room temperature. This was followed by 3 hours of dialysis against 10 mM HEPES, 1 mM DTT pH 7.5 and an over night dialysis against 10 mM HEPES pH 7.5 at 4 °C. The sample was then loaded on the Superdex G75 HiLoad 16/60 gel filtration column and eluted fractions were analysed by Western blot. Protein concentration in the enzyme preparation was measured spectrophotometrically at 280 nm using the theoretically calculated molar extinction coefficient (ε =28810 M^{-1} cm^{-1}). Peptide mass finger-printing and sequencing was conducted to confirm the identity of excised bands from the SDS-PAGE gel of the produced protein.

7.1.5 GWD3-SBD *P. pastoris* EndoH treatment

Endoglycosidase H (EndoH) is a recombinant glycosidase which cleaves mannose and some oligosaccharides from N-linked glycoproteins. 1200 µL GWD3-SBD [1.33 mg/mL] added 75 µL endoH (New England Biolabs), 130 µL 10x NEB buffer and 56 µL protease inhibitor (25 x conc stock) were mixed and left 20 hours at room temp. The pH was adjusted by to 7.5 by 15 µL NaOH (0.5 M) and 200 µL glycerol and 190 µL NaCl (5M) were added to adjust conditions for binding to a NiNTa spin column (QIAgen). The column were equilibrated with binding buffer (20 mM $NaPO_4$, 20 mM imidazole, 500 mM NaCl, 5 mM DTT and 10% glycerol pH 7.4) and 500 µL sample was loaded and washed twice with binding buffer. Samples were eluted by adding 200 µL elution buffer (20 mM $NaPO_4$, 500 mM imidazole, 500 mM NaCL, 5 mM DTT and 10% glycerol pH 7.4).

7.1.6 GA-SBD *E. coli* Gateway

Plasmid construction

A plasmid containing the full-length *A. niger* GA cDNA (SwissProt accession no. P69328) was used as a PCR template. Phusion™ High-Fidelity DNA Polymerase (New England Biolabs) was used to amplify Cys^{509}-Arg^{616} with primers P16 and

P17 (Appendix table 9.1). The *attB* PCR product was subsequently by recombination cloned into pDONR201 vector by the Gateway BP recombination reaction according to manufacturer's instructions (Invitrogen). The insert from the Gateway entry clone was transferred by recombination into pDEST14 and pDEST15 to make the Gateway expression plasmids. The plasmids were electroporated into *E. coli* BL21A1 (Invitrogen) for expression and into *E .coli* TOP10 (Invitrogen) for long time storage and correctly inserted DNA was verified by sequencing.

Expression and purification

E. coli strain BL21A1 harbouring pDEST14 with GA-SBD were grown at 37 ºC in 1 L LB medium supplemented with 100 µg/mL ampicillin to and optical density of 0.4 at 600 nm. L-arabinose was added to a final concentration of 0.2% and the cultivation was continued at 28 ºC and 37 ºC. Samples were collected every hour to verify expression by SDS-PAGE. No overexpression could be seen at all on the SDS-PAGE and therefore work on this construct was abandoned to try other expression vectors.

E. coli strain BL21A1 harbouring pDEST15 with GA-SBD were grown as described above. Antibodies against the GST tag were used to confirm GA-GST fusion expression. Cells grown for 5 hours at 28 ºC were pelleted by centrifugation at 5000 x g for 20 min at 4 ºC. After harvest the cells were resuspended in 10 mL (500 mM Tris-HCl, 30 mM EDTA, 0.5 mM PMSF, 5 mM DTT pH 7) and disrupted by sonication. The extracts were clarified by centrifugation 12,000 x g for 30 min at 4 ºC. The samples were loaded with a flow of 0.2 mL/min on a 1 mL GSTrap-column (GE-Healthcare) and washed with 20 CV of washing buffer (140 mM NaCl, 2,7 mM KCl, 10 mM Na_2HPO_4, 1.8 KH_2PO_4, 5 mM DTT pH 7.3) with a flow of 0.7 mL/min and eluted with a gradient of 10 CV in elution buffer (50 mM Tris, 20 mM GSH, 200 mM NaCl and 10% glycerol, pH 8.0) and a flow of 0.5 mL/min. Eluted fractions were analysed by SDS-PAGE.

7.1.7 GA-SBD *P. pastoris* pPICZα B

Plasmid construction

A plasmid containing the full-length *A. niger* GA cDNA (SwissProt accession no. P69328) was used as a PCR template and the GA-SBD construct was amplified using primer P18 and P19 (Appendix table 9.1). An aliquot of the PCR product was digested with *Sfi*I and *Xba*I and ligated into pPICZα B digested with the same enzymes. The plasmids were transformed into DH5α by electroporation. Positive transformants were selected and propagated on low salt LB plates with 25 µg/mL zeocin and propagated. Plasmids were purified using miniprep (QIAgene) and sequenced. After sequencing the plasmids with correct inserted DNA were linearized by *Pme*I and transformed into *P. pastoris* strain X33 by electroporation. Transformants were selected on YPDS plates (1% (w/v) yeast extract, 2% (w/v) peptone, 2% (w/v) glucose, 1 M sorbitol, 2% (w/v) agar and 100 µg/mL zeocin). Transformants secreting GWD3-SBD were screened on MM plates (1.34% (w/v) yeast nitrogen base, 4 x 10^{-5}% (w/v) biotin, 0.5% (v/v) methanol and transferred

onto a nitrocellulose membrane using replica plating. Positive transformants were detected by Western blotting using the c-myc antibody.

7.1.8 LAF-SBD *E. coli* Gateway

Plasmid construction

A plasmid containing the full-length *Homo sapiens* Laforin cDNA (SwissProt accession no. O95278.) was used as a PCR template. Phusion™ High-Fidelity DNA Polymerase (New England Biolabs) was used to amplify Met^{118}-Glu^{251} with primer P20 and P21 and Glu^{138}-Glu^{251} with primer P22 and P23 (Appendix table 9.1). The *attB* PCR product was subsequently by recombination cloned into pDONR201 vector by the Gateway BP reaction according to manufacturer's instructions (Invitrogen). The insert from the Gateway entry clone was transferred by recombination into pDEST15 to make the Gateway expression plasmids. The plasmids were electroporated into *E. coli* BL21A1 (Invitrogen) for expression and into *E .coli* TOP10 (Invitrogen) for long time storage and correct inserted DNA was verified by sequencing.

Expression and purification

E. coli strain BL21A1 harbouring pDEST15 with LAF-SBD Met^{118}-Glu^{251} and LAF-SBD Glu^{138}-Glu^{251} were grown at 37 ºC in 1 L LB medium supplemented with 100 µg/mL ampicillin to OD_{600} ~0.4. L-arabinose was added to a final concentration of 0.2% and the cultivation was continued at 28 ºC. Samples were collected every hour to verify expression by SDS-PAGE. Antibodies against the GST tag were used to confirm LAF-GST fusion expression. Cells grown for 3 hours at 28 ºC were pelleted by centrifugation at 5000 x g for 20 min at 4 ºC. After harvest the cells were resuspended in 10 mL (500 mM Tris-HCl, 30 mM EDTA, 0.5 mM PMSF, 5 mM DTT pH 7) and disrupted by sonication. The extracts were clarified by centrifugation 12000 x g for 30 min at 4 ºC.

To purify LAF-SBD the clarified samples were loaded with a flow of 0.2 mL/min on a 1 mL GSTrap-column (GE-Healthcare) and washed with 20 CV of washing buffer (140 mM NaCl, 2,7 mM KCl, 10 mM Na_2HPO_4, 1.8 KH_2PO_4, 5 mM DTT pH 7.3) with a flow of 0.7 mL/min and eluted with a gradient of 10 CV in elution buffer (50 mM Tris, 20 mM GSH, 200 mM NaCl and 10% glycerol, pH 8.0) and a flow of 0.5 mL/min. Eluted fractions were analysed by SDS-PAGE.

7.1.9 GST control expression and purification

To analyse the carbohydrate interacting effect of the GST tag alone a pET41a(+) plasmid (Novagen) was used to express the GST protein alone.

BL21A1 cells harbouring pET41a(+) were grown at 37 ºC in 1 L LB medium supplemented with 200 µg/mL kanamycin to and optical density of 0.5 at 600 nm. IPTG was added to a final concentration of 0.5 mM and the cultivation was continued at 28 ºC. Samples were collected every hour to verify expression by SDS-PAGE. Cells grown for 5 hours at 28 ºC were pelleted by centrifugation at 5000 x g for 20 min at 4 ºC. After harvest the cells were resuspended in 10 mL (500 mM Tris-HCl, 30 mM EDTA, 0.5 mM PMSF and 5 mM DTT, pH 7) and disrupted by

sonication. The extracts were clarified by centrifugation 12000 x g for 30 min at 4 ºC.
The samples were loaded with a flow of 0.2 mL/min on a 1 mL GSTrap-column (GE-Healthcare) and washed with 20 CV of washing buffer (140 mM NaCl, 2,7 mM KCl, 10 mM Na_2HPO_4, 1.8 KH_2PO_4, 5 mM DTT pH 7.3) with a flow of 1 mL/min and eluted in a step elution with elution buffer (50 mM Tris, 20 mM GSH, 200 mM NaCl and 10% glycerol, pH 8.0). Eluted fractions were analysed by SDS-PAGE and were only ~80% pure and therefore eluted fractions were mixed and dialysed against 20 mM HEPES pH 7.5 to prepare the sample for HisTrap purification. After dialysis the sample was added 500 mM NaCl and 10 mM imidazole and loaded on a 1 mL HisTrap column equilibrated in washing buffer (20 mM HEPES, 20 mM imidazole, 500 mM NaCl, pH 7.4). Bound protein was washed with a flow of 1 mL/min 20 CV and eluted with a flow of 0.5 mL/min in elution buffer (20 mM HEPES, 500 mM imidazole, 500 mM NaCl, pH 7.4). Eluted fractions were analysed by SDS-PAGE.

7.1.10 GA *P. pastoris* pPICZα B

Plasmid construction

A plasmid containing the full-length *A. niger* GA cDNA (SwissProt accession no. P69328) was used as a PCR template and primer P23 and P19 (Appendix table 9.1) was used to amplify the *P. pastoris* GA construct. An aliquot of the PCR product was digested with *Sfi*I and *Xba*I and ligated into pPICZα B digested with the same enzymes. The plasmids were transformed into DH5α by electroporation. Positive transformants were selected on low salt LB plates with 25 µg/mL zeocin and propagated. Plasmids were purified using miniprep (QIAgene) and sequenced. After sequencing the plasmids with correct inserted DNA were linearized by *Pme*I and transformed into *P. pastoris* strain X33 by electroporation. Transformants were selected on YPDS plates (1% (w/v) yeast extract, 2% (w/v) peptone, 2% (w/v) glucose, 1 M sorbitol, 2% (w/v) agar and 100 µg/mL zeocin). Transformants secreting GA were screened on MM-starch plates (1.34% (w/v) yeast nitrogen base, 4 x 10^{-5}% (w/v) biotin, 0.5% (v/v) methanol, 2% (w/v) soluble potato starch) and transferred onto a nitrocellulose membrane using replica plating. Positive transformants were detected by Western blotting using the c-myc antibody.

Expression and purification

Transformants were grown in 0.5 L BMGY (1% (w/v) yeast extract, 2% (w/v) peptone, 100 mM potassium phosphate, pH 6.0, 1.34% (w/v) yeast nitrogen base, 4 x 10^{-5}% (w/v) biotin, 1% (v/v) glycerol) at 30˚C for 18 hours in 3 L Erlenmeyer flasks to OD_{600} ~ 2. Cells were harvested (20 ˚C, 1500 rpm, 8 min) and resuspended in 1L BMMY (1% (w/v) yeast extract, 2% (w/v) peptone, 100 mM potassium phosphate, pH 6.0, 1.34% (w/v) YNB, 4 x 10^{-5}% (w/v) biotin and 0.5% (v/v) methanol) to induce expression. The culture was incubated for 22 hours at 30˚C shaken with 200 rpm. Cells were harvested by centrifugation (4 ˚C, 9000 rpm, 20 min) and the supernatant was collected. Samples from different induction time points were analysed by Western blot and expression of GA was verified by antibodies raised against GA and the c-myc epitope.

To prepare the sample for HisTrap purification the pH of the supernatant was adjusted to 7.4 by NaOH and 10 mM imidazole, 500 mM NaCl, were added and after stirring the supernatant was filtered through 0.22 μm filter. The supernatant was loaded on a 1 mL HisTrap column (QIAgen) and a flow of 1 mL/min, washed with 15 column volumes and eluted with a gradient of 20 CV from 0 - 100 % elution buffer (20 mM sodium phosphate, 0.5 M NaCl, 500 mM imidazole, pH 7.5). Eluted protein was analysed by SDS-PAGE.

7.1.11 GA without SBD *P. pastoris* pPICZα B

Plasmid construction

A plasmid containing the full-length GA cDNA (SwissProt accession no. P69328) was used as a PCR template. Phusion™ High-Fidelity DNA Polymerase (New England Biolabs) was used to amplify amino acid sequence (1-508) with primer P23 and P25 (Appendix table 9.1). An aliquot of the PCR product was digested with *Sfi*I and *Xba*I and ligated into pPICZα B digested with the same enzymes. The plasmids were transformed into DH5α by electroporation. Positive transformants were selected on low salt LB plates with 25 μg/mL zeocin and propagated. Plasmids were purified using miniprep (QIAgene) and sequenced. After sequencing the plasmids with correct inserted DNA were linearized by *Pme*I and transformed into *P. pastoris* strain X33 by electroporation. Transformants were selected on YPDS plates (1% (w/v) yeast extract, 2% (w/v) peptone, 2 % (w/v) glucose, 1 M sorbitol, 2% (w/v) agar and 100 μg/mL zeocin). Transformants secreting GA without SBD were screened on MM plates (1.34% (w/v) yeast nitrogen base, 4 x 10^{-5}% (w/v) biotin, 0.5% (v/v) methanol and transferred onto a nitrocellulose membrane using replica plating. Positive transformants were detected by Western blotting using the c-myc antibody and antibodies raised against GA.

Expression and purification

Transformants were grown in 0.5 L BMGY (1% (w/v) yeast extract, 2% (w/v) peptone, 100 mM potassium phosphate, pH 6.0, 1.34% (w/v) yeast nitrogen base, 4 x 10^{-5}% (w/v) biotin, 1% (v/v) glycerol) at 30˚C for 18 hours in 3 L Erlenmeyer flasks to OD_{600} ~ 2. Cells were harvested (20˚C, 1500 rpm, 8 min) and resuspended in 1L BMMY (1% (w/v) yeast extract, 2% (w/v) peptone, 100 mM potassium phosphate, pH 6.0, 1.34% (w/v) YNB, 4 x 10^{-5}% (w/v) biotin and 0.5% (v/v) methanol) to induce expression. The culture was incubated for 22 hours at 30 ˚C shaken with 200 rpm. Cells were harvested by centrifugation (4˚C, 9000 rpm, 20 min) and the supernatant was collected. Samples from different induction time points were analysed by Western blot and expression of GA was verified by antibodies raised against GA or the c-myc epitope.

The pH of the supernatant was adjusted to 7.4 by NaOH and 10 mM imidazole, 500 mM NaCl, were added to prepare the sample for HisTrap purification and after stirring the supernatant were filtered trough 0.22 μm filter. The supernatant were loaded on a 1 mL HisTrap column (QIAgen) and a flow of 1 mL/min, washed with 15 column volumes and eluted with a gradient of 20 CV from 0 - 100% elution buffer (20 mM sodium phosphate, 0.5 M NaCl, 500 mM imidazole, pH 7.5). Eluted protein was analysed by SDS-PAGE.

7.1.12 GA-GWD3 fusion *P. pastoris* pPICZα B

Plasmid construction

The plasmid encoding the GA-GWD3 fusion in *P. pastoris* was constructed by PCR amplification in two reactions. First the GA part without SBD and with an overlap sequence was amplified by primer P23 and P26 and the GWD3-SBD part with an overlap sequence was amplified by primer P15 and P25 (Appendix table 9.1). In the second reaction the two amplified fragments from reaction one was mixed and the following PCR programme was used:

Step 1	95 ˚C	2 min
Step 2	95 ˚C	40 sec
Step 3	60 ˚C	30 sec
Step 4	72 ˚C	2 min
Step 5	72 ˚C	7 min

Primers P23 and P15 were added to amplify the fusion after two rounds of cycle 2-4. Cycle 2- were repeated 30 times in total. An aliquot of the PCR product was digested with *Sfi*I and *Xba*I and ligated into pPICZα B digested with the same enzymes. The plasmids were transformed into DH5α by electroporation. Positive transformants were selected on low salt LB plates with 25 µg/mL zeocin and propagated. Plasmids were purified using miniprep (QIAgene) and sequenced. After sequencing the plasmids with correct inserted DNA were linearized by *Pme*I and transformed into *P. pastoris* strain X33 by electroporation. Transformants were selected on YPDS plates (1% (w/v) yeast extract, 2% (w/v) peptone, 2% (w/v) glucose, 1 M sorbitol, 2% (w/v) agar and 100 µg/mL zeocin). Transformants secreting GA-GWD3 fusion were screened on MM-starch plates (1.34% (w/v) yeast nitrogen base, 4 x 10^{-5}% (w/v) biotin, 0.5% (v/v) methanol, 2% (w/v) soluble potato starch) and transferred onto a nitrocellulose membrane using replica plating. Positive transformants were detected by Western blotting using the c-myc antibody.

Expression and purification

Small scale pilot expression tests were made to optimise expression. Different temperatures during induction, different methanol concentrations for induction, addition of casaminoacids and different shake flask options were evaluated and to determine the optimal harvest time, and samples were collected from different time points during induction and analysed by SDS-PAGE or Western blot. Optimal expression conditions were as following: transformants were grown in 0.5 L BMGY (1% (w/v) yeast extract, 2% (w/v) peptone, 100 mM potassium phosphate, pH 6.0, 1.34% (w/v) yeast nitrogen base, 4 x 10^{-5}% (w/v) biotin, 1% (v/v) glycerol) at 30 ˚C for 18 hours in 3 L Erlenmeyer flasks to OD_{600} ~ 2. Cells were harvested (20 ˚C, 1500 rpm, 8 min) and resuspended in 1L BMMY (1% (w/v) yeast extract, 2% (w/v) peptone, 100 mM potassium phosphate, pH 6.0, 1.34% (w/v) YNB, 4 x 10^{-5}% (w/v) biotin and 0.5% (v/v) methanol) to induce expression. The culture was incubated for 22 hours at 30 ˚C shaken with 200 rpm. Cells were harvested by centrifugation (4˚C, 9000 rpm, 20 min) and the supernatants were collected. Samples from different induction time points were analysed by Western blot and

expression of GA-GWD3 fusion was verified by antibodies raised against GA or the c-myc epitope.
The pH of the supernatant was adjusted to 7.4 by NaOH and 10 mM imidazole, 500 mM NaCl, were added and after stirring the supernatant was filtered trough 0.22 μm filter. The supernatant were loaded on a 1 mL HisTrap column (QIAgen) and a flow of 1 mL/min, washed with 15 column volumes and eluted with a gradient of 20 CV from 0 - 100% elution buffer (20 mM sodium phosphate, 0.5 M NaCl, 500 mM imidazole, pH 7.5). Eluted protein was analysed by SDS-PAGE and fractions with GA-GWD3 fusion were combined. The combined fractions from the NiNTA column were dialysed against a 10 mM BICINE buffer pH 8.5 and loaded on a 6 mL Resource Q column (GE-Healthcare) equilibrated in the same buffer with a flow of 0.5 mL/min. The column was washed with 10 CV and a flow of 2mL/min and eluted with a gradient of 18 CV and a flow of 0.5 mL/min with 10 mM BICINE, 500 mM NaCl buffer pH 8.5. Fractions of bound protein was analysed on SDS-PAGE.
The eluted fractions were combined and dialysed against a 10 mM sodium acetate buffer pH 5.5 and loaded on a 6 mL Resource S column (GE-Healthcare) equilibrated in the same buffer with a flow of 0.5 mL/min. The column was washed with 10 CV and a flow of 2 mL/min and eluted with a gradient of 10 CV and a flow of 0.5 /min with 10 mM sodium acetate, 500 mM NaCl buffer pH 5.5. Eluted fractions was analysed by SDS-PAGE. The eluted fractions were combined and dialysed against 20 mM sodium acetate pH 5.5 and after dialysis the sample was supplemented with $(NH_4)_2SO_4$ up to 5% (w/v) and the sample was applied to a 6 mL β-CD linked Sepharose column equilibrated with binding buffer (20 mM sodium acetate, 250 mM NaCl pH 5.5). The column was washed with 2 column volumes binding buffer and eluted in 20 mM sodium acetate, 10 mM β-CD, pH 5.5. Eluted fractions were concentrated (Centriprep, 30 kDa cut-off, Millipore) and analysed by SDS-PAGE. The GA-GWD3 fusion did not bind to the β-CD linked Sepharose column and was found in the flow through fraction.
The flow through from the β-CD linked Sepharose column was dialysed against 10 mM HEPES pH 7.5 and loaded with a flow of 0.7 mL/min on a HiLoad 16/60 Superdex G75 gel filtration column equilibrated with 10 mM HEPES, 200 mM NaCl pH 7.5 and eluted peaks were analysed by SDS-PAGE.
Eluted protein was stored at 4 ºC. Protein concentration was measured spectrophotometrically at 280 nm using the theoretically calculated molar extinction coefficient (ε =141915 M^{-1} cm^{-1}). SDS-PAGE analysis was carried out to assess purity. Peptide mass finger-printing was conducted to identity degradation products.

7.2 Homology modelling

A three dimensional model of GWD3-SBD was built by homology modelling using Modeler v6.1 (Sali and Blundell, 1993). The SBD of GA from *A. niger* (PDB code 1KUL, 26% identity and 57.7% similarity) residues 509-616 was used as template. Residues 170-184 from the GWD3-SBD were deleted from the model structure, since there were no corresponding residues in the *A. niger* GA-SBD. Because of the low homology to the template, the structure was relaxed by molecular dynamics simulation, using the program NAMD (Kalé *et al.,* 1999). To build a homolog

model of the *H. sapiens* LAF-SBD the *B. circulans* strain 251 CGTase (PDB code 2DIJ, 19% identity and 31% similarity) residues 584-686 was used as template using Modeler v6.1 (Sali and Blundell, 1993).

7.3 Differential scanning calorimetry (DSC)

The stability of the SBD was assessed by differential scanning calorimetry using a VP-DSC calorimeter (MicroCal, Northampton, MA) with a cell volume of 0.52 mL (Plotnikov *et al.,* 1997). Protein solutions with a concentration of 2 mg/mL in 50 mM HEPES, pH 8.0 were used. Protein samples were spun down at 4 °C for 40 min at 14000 g and sterile filtered prior to the run. Buffer from the last dialysis step was sterile filtered and used to collect blank scans. Samples and buffers were degassed by vacuum at 15 °C for 15 min. All scans were collected between 15 to 100 °C at a rate of 1 °C/min. Samples were pre-temperated for 10 min at 15 °C. In order to remove residual β-CD present after protein purification, purified protein was subjected to an additional His-trap purification step and rescanned as above following buffer exchange.

7.4 Surface Plasmon Resonance (SPR)

SPR was performed using a Biacore T100 (GE healthcare). The SBDs were biotinylated in the presence of 10 mM β-CD (GWD3-SBD) or 1 mM β-CD (GA-SBD) to minimise biotinylation at site that would block the carbohydrate binding sites, and in the presence of EZ-Link1 Sulfo-NHS-LC-Biotin (sulfosuccinimidyl-6-(biotinamido) hexanoate) (Pierce). The reagent was added to 100 µL protein (70-100 µM, in 50 mM HEPES pH 7.5) to a final 10-fold molar excess in a total volume of 120 µL. The reaction mixture was incubated for 30 min at 20 ºC and labelled protein was desalted twice on a 0.5 mL Zeba Desalt Spin Column (Pierce). Streptavidin-coated sensor chips (Sensor Chip SA) were preconditioned by injecting 30 µL 1M NaCl, 20 mM NaOH at a flow rate of 30 µL/min three times and biotinylated protein 5 ng/µL were immobilized was injected at a flow rate of 10 µL/min. Carbohydrate ligands were dissolved in running buffer (20 mM HEPES pH 7.0, 20 mM sodium acetate pH 5.5 or 20 mM Tris-HCl pH 9.0) including 0.5% P20 surfactant (GE healthcare). Binding of α-, β- and γ-CD was analysed at 15 different concentrations in the range 0.01−10 mM. α-and γ-CD were analysed in duplicates and β-CD using four replicates. Binding experiments were performed at 10 °C both in the sample compartment and in the SPR cells at a flow rate of 30 µL/min. Data from parallel flow cells without enzyme served as reference and were subtracted from sample sensorgrams. The apparent equilibrium dissociation constant (K_d), was determined by using steady state affinity fitting analysis (equation 1) assuming 1 binding site (BIAevalution version 1.1 software (GE healthcare)) and Sigma Plot 9.0.

$$R = \frac{R_{\max} \cdot [\beta - CD]}{[\beta - CD] + K_d} \qquad \text{Eq 1}$$

where R is the difference in response between parallel flow cells with and without protein, [β-CD] is the concentration of β-CD, K_d is the equilibrium dissociation constant and R_{max} is the maximum binding capacity. Examine the binding data assuming two independent binding sites equation 2 was used.

$$R = \frac{R_{\max 1} \cdot [\beta - CD]}{[\beta - CD] + K_{d1}} + \frac{R_{\max 2} \cdot [\beta - CD]}{[\beta - CD] + K_{d2}} \qquad \text{Eq 2}$$

where R, β-CD and K_d are as in eq. 4.

7.5 Confocal Laser Scanning Microscopy (CLSM)

Proteins were labelled with fluorescein 5-EX succinimidyl ester (Invitrogen) by amine modification. Fluorescein 5-EX succinimidyl ester was dissolved in dimethyl formamide at 10 mg/mL immediately before starting the reaction. The reagent was added to 100 µL GWD3-SBD (1 mg/mL, in a 50 mM HEPES, 10 mM β-CD buffer adjusted to pH 8 by 0.2 M sodium carbonate) in a final 4-fold molar excess in a total volume of 140 µL. The reaction mixture was incubated for 1 hour at 20 ºC with continuous stirring.

The fluorescein conjugated proteins were separated from excess reagent by desalting twice on a 0.5 mL Zeba Desalt Spin Column (Pierce). Labelled protein binding to starch granules was visualised by confocal laser scanning microscopy (TCS SP2, Leica Microsystems, Germany) as previously described (Blennow *et al.,* 2003). For fluorescein, a 488 nm laser was used for excitation, and emission of light was detected between 500 and 550 nm. A laser power of 25% was maintained during acquisition of all images, and the gain was varied to prevent saturation of the detector and to ensure comparable intensities. The objectives used were PL APO CS 20.0x0.70 and PL APO 63.0x1.20 water immersion objectives. Image analysis was performed with the TCS SP2 software.

7.6 Transient expression of GWD3-SBD in Tobacco

A fragment containing the 201 N-terminal amino acids of GWD3 was PCR amplified using uracil-containing primers and cloned into the vector pPS48sYFP using an improved USER™ (uracil-specific excisionreagant, New England Biolabs) cloning procedure (Nour-Eldin *et al.,* 2006). The construct was transformed into *Agrobacterium tumefaciens* strain C58C1 and was grown at 29 °C in LB supplemented with 50 µg mL^{-1} kanamycin and 5 µg mL^{-1} tetracycline to stationary phase. Bacteria were pelleted by centrifugation at 5000 g for 15 min at room temperature and resuspended in 10 mM $MgCl_2$ and 150 µg mL^{-1} acetosyringone. Cells were left in this medium for 3h and then infiltrated into 2-4 week-old *Nicotiana benthamiana* plants as described in (Voinnet *et al.,* 2003). Leaves transiently expressing the YFP-GWD3-SBD fusion were analysed by CLSM (TCS SP2, Leica Microsystems, Germany) equipped with a 63x1.20 PL APO water immersion objective. A 488 nm laser line was used for excitation and emission was detected between 520 nm and 550 nm for YFP fluorescence and 650 nm and 750

nm for chlorophyll auto-fluorescence. Image analysis was performed with the TCS SP2 software.

7.7 Starch granule binding

Protein (5-10 nM) (6 μg) was incubated at 4 °C with various types of starch granules (barley, maize, potato and waxy maize) in 20 mM HEPES 0.005% (w/v) bovine serum albumin, mixed for 45 min and centrifuged (4000g, 4 °C, 5 min). The experiments were performed with 6 concentrations of starch granules (0-150 mg/mL). Supernatants were transferred to new Eppendorf tubes and amount of unbound protein were detected by several methods. GST-tagged proteins were analysed by GST assay and by Western blot and detection by GST antibody. Proteins without GST tag were analysed by fluorescence spectrophotometer, or measured spectrophotometrically at 280 nm using the theoretically calculated molar extinction coefficient or by SDS-PAGE.

8 Results

8.1 Production of recombinant proteins

8.1.1 GWD3-SBD

***E. coli* Gateway GWD3-SBD**

The length of GWD3-SBD was defined by using the *A. niger* glucoamylase SBD as template and resulted in a 102 amino acid long polypeptide (amino acid residues 68-170, Fig. 7.1) (Baunsgaard *et al.,* 2005). This sequence was used to create an expression vector to express the GWD3-SBD without any tag in *E. coli* (pDEST14). Small scale expressions were performed with an induction temp of 28 °C and 37 °C and samples were collected from 1 hour until 20 hours after induction. The samples were analysed on SDS-PAGE but there were no sign of expressed GWD3-SBD. There were no available antibodies against the GWD3-SBD and therefore another expression vector was chosen. The pDEST15 was chosen for several reasons. The GST tag is three times bigger than the GWD3-SBD and could mimic the catalytic domain of GWD3 and thereby stabilise the SBD and also the possibility of exploiting the GST tag to test activity in starch binding assays and finally accessibility of the commercial available antibodies against GST. An expression vector with the amino acid sequences 68-170 was created with an N-terminal GST tag. Expression of GWD3-SBD in pDEST15 was analysed at two different induction temperatures 28 °C and 37 °C and the highest expression level was found after 5 hours induction at 28 °C. GWD3-SBD was captured by GST-tag affinity chromatography which resulted in 40-50% pure fractions based on SDS-PAGE. The purification was optimised by using a gradient elution and by adding 10% glycerol to the buffers and resulted in 70-80% pure GWD3-SBD (Fig. 5.1: A lane1). Binding of GWD3-SBD from pDEST15 to β-CD linked Sepharose was not successful as the fusion protein did not bind the resin. Anion exchange purification resulted in 80-85% pure protein, but at the same time also an increase in two lower molecular weight bands was observed (Fig. 8.1A, lane 2). The size of the upper band two corresponds to the size of the GST tag alone (27.7 kDa) and the low molecular mass band could be a degradation product of the GWD3-SBD (~8 kDa) and both proteins are undesirable for further investigations of the SBD. A gel filtration step was applied to improve purity, but also after this purification step the same bands appeared and the ratio of GWD3-SBD to GST tag alone was reduced. To obtain more pure protein the same amino acid sequence 68-170, as in the pDEST15 vector, was used to create a pDEST17 expression vector with an N-terminal 6 x His tag. Same expression conditions as applied to the pDEST15 vector was used for the pDES17 vector and the cells were harvested after 5 hours induction. A clear relative increase in the intensity of one band could be seen with a molecular weight ~14 kDa (Fig 8.1, B Lane 1), which is in reasonable agreement with the expected (14.6 kDa). Affinity chromatography was used to purify GWD3-SBD pDEST17 and after HisTrap purification the samples were analysed by SDS-PAGE. Proteins of the correct size was found in the flow through and in the wash and very little seemed to

bind to the column (Fig 8.1, B). A 1 mL NTA column was also tested but with the same result. To analyse if the lack of binding to the HisTrap column and the slightly smaller than expected size was a result of a cleavage of the 6xHis tag the GWD3-SBD band was cut out of the gel and prepared for mass spectrometry analysis. No peptide was found with the size corresponding to the N-terminal part of the protein verifying lack of the N-terminal His tag in the expressed protein.

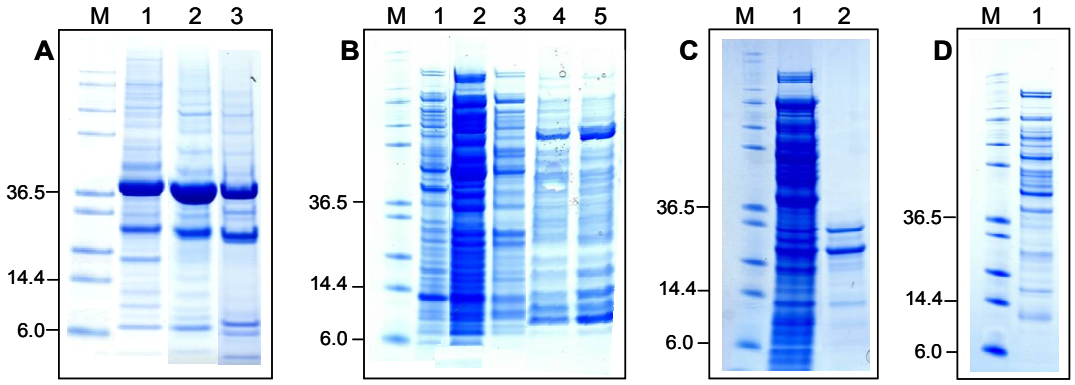


Figure 8.1: Purification of *E. coli* pDEST15 and 17, GWD3-SBD, analysed by SDS-PAGE.
A: Lane 1: GWD3-SBD pDEST15 after GSTrap purification, Lane 2: GWD3-SBD pDEST15 after purification on a HiTrap Q FF column. Lane 3: GWD3-SBD pDEST15 after P30 gel filtration purification. **B** Lane 1: GWD3-SBD pDEST17 before loading on a 1 mL HisTrap column. Lane 2: Flow through. Lane 3: Wash Lane 4: Elution from first peak. Lane 5: Elution from second peak. **C** Lane 1: GWD3-SBD pDEST17 before loading on a HiLoad superdex 75 (16/60) gel filtration column. Lane 2: Elution from gel filtration column. **D** Lane 1: GWD3-SBD pDEST17 after elution from a Resource Q column. M: molecular weight standard.

The GWD3-SBD pDEST17 did not bind to a β-CD linked Sepharose column tested at two different pH values (5.5 and 7.5) and instead gel filtration purification was tested. The GWD3-SBD pDEST17 was found in a diluted concentration and at least two other proteins with higher relative amounts than the GWD3-SBD pDEST17 were eluted in the same fraction and more strongly presented than the GWD3-SBD pDEST17 (Fig.8.1, C). The amount of GWD3-SBD pDEST17 did not represent the amount of loaded protein and this suggested the protein was degraded during the purification step. Also purification by anion-exchange did not improve purity to a satisfactory level (Fig. 8.1, D) and therefore the conclusion was that the protein was not properly folded or stable enough and another expression system was chosen to examine if the protein behaved differently when expressed in *Pichia pastoris*.

P. pastoris GWD3-SBD

The amino acid sequence (68-170) was cloned into the pPICZα B vector and transformants expressing GWD3-SBD was verified by colony screening and 10 positive transformants were selected for further expression tests. The theoretically molecular mass of the GWD3-SBD construct in the pPICZα vector is 13.9 kDa, but since there are two possible N-glycosylations sites an apparent molecular mass higher than this would be expected if the enzyme was modified in *P. pastoris*. After

small scale expression tests one transformant was selected and highest level of expression was found after 18 hours at 20 °C (Fig. 8.2, A). After 18 hours of induction the level of secreted protein declined and instead the amount of a smaller degradation product increased. The smaller degradation product was recognised by the c-myc antibody and the since the c-myc epitope is located at the C-terminal end of the protein the degradation must appear in the N-terminal part and the represent a loss of approximately 4 kDa.
It also became clear (Fig. 8.2 A, Lane 5 and 6) that the domain seems to dimerise most likely because of intramolecular disulfide bond formation between some of the cysteines found in the domain. After an affinity chromatography purification there was still trace of other proteins and the degradation products recognised by the c-myc antibody also bound to the HisTrap column and were eluted from the column with the same imidazole concentration. An intense smearing (possibly due to in gel-aggregation or from the glycosylation of the *Pichia* GWD3-SBD) was reduced when samples were eluted with high glycerol concentration in the elution buffer (20%).

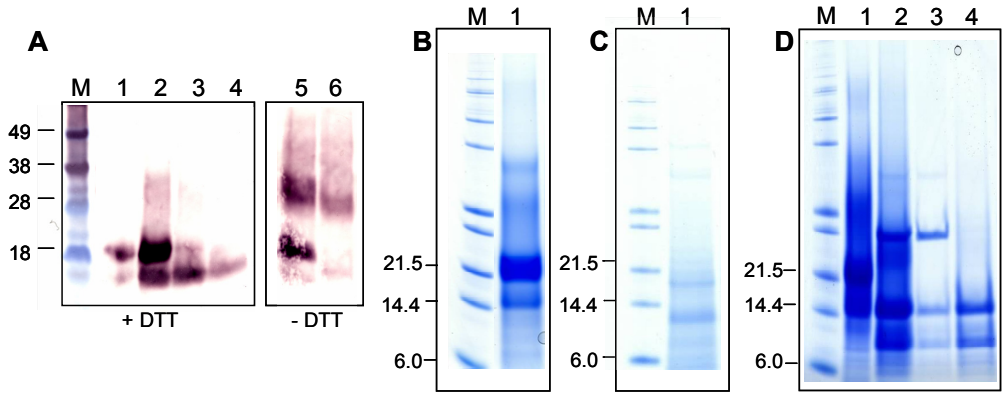


Figure 8.2: Expression and purification of *P. pastoris* GWD3-SBD
A: Western Blot of GWD3-SBD expression. Lane 1: 6 hours after induction. Lane 2: 18 hours after induction. Lane 3: 30 hours after induction. Lane 4: 48 hours after induction. Lane 5: 18 hours after induction. Lane 6: 30 hours after induction. Lane 1-4 was analysed under reducing conditions and lane 5-6 under non-reducing conditions. **B**-**D** SDS-PAGE of GWD3-SBD purification and EndoH treatment. **B**: Lane 1: elution from 1 mL HisTrap column. **C** Elution from gel filtration column after reduction of the sample before loading. **D**: GWD3-SBD after 1 mL HisTrap column and before EndoH treatment. Lane 2: GWD3-SBD after EndoH treatment. Lane 3: Flow through from the second HisTrap purification after EndoH treatment. Lane 4: purified EndoH treated GWD-SBD. M: molecular weight standard.

Despite many attempts to purify the *P. pastoris* expressed GWD3-SBD by ion exchange and gel filtration, no pure protein was obtained. During elution from the gel filtration column samples containing GWD3-SBD eluted very early indicating dimerisation or oligomerisation. A new sample, which was reduced with DTT before applying to the column, was prepared and this reduction shifted the elution time point of the *Pichia* GWD3-SBD towards a smaller protein size. The sample was a mixture of the GWD3-SBD and the degradation product and with a decrease in amount of *Pichia* GWD3-SBD compared to the degradation product and still

with some higher molecular weight bands contaminating the sample. The sample was also loaded on a β-CD linked Sepharose column but the *Pichia* GWD3-SBD did not bind to the matrix at either pH 5.5 or pH 7.5. During storage at 5 °C the *Pichia* GWD3-SBD was degraded and after two weeks most of this protein was degraded to the 10 kDa fragment and also this degradation product was further degraded during storage.
The *Pichia* GWD3-SBD was EndoH treated to analyse the purity of protein fraction and the stability after removal of the glycosylation (Fig. 8.2, D). After removal of the *N*-glycosylation the *Pichia* GWD3-SBD appeared on an SDS-PAGE gel with the expected size 13.9 kDa and the degradation product also seemed to be affected by the EndoH treatment and a decrease in size was observed (~10 kDa). Mass spectrometric analysis and N-terminal sequencing of the two bands confirmed that both bands contained sequences from the *Pichia* GWD3-SBD and that the upper band had the correct N-terminal sequence. The lower band seemed to be a mixture of degradation products with no clear degradation site.

***E.coli* GWD3-SBD with longer CBM border at the C-terminus**

The GWD3 amino acid sequence (68-170) tagged by a C- or N-terminal His tag or by a GST tag, resulted in low production yields in *E. coli* and *P. pastoris* and despite extensive efforts to purify the GWD3-SBD by conventional chromatographic methods such as ion exchange, affinity and gel filtration chromatographic, highly pure protein was not obtained. Furthermore, stability was very low and the proteins were degraded within weeks. Further analysis of the primary structure suggested that the GWD3-SBD might have a longer C-terminus as judged by the presence of hydrophobic residues downstream of the annotated sequence and the lack of an apparent secondary structure shift such as the presence of glycines or prolines. Two additional constructs encoding amino acid sequence 68-179 and 68-184 (Fig. 7.1) were prepared in *E. coli* fused to a C-terminal His tag.
These two constructs were expressed at an induction temperature of 17 °C and after 20 hours the cells were not growing anymore. Analysis by SDS–PAGE of soluble and insoluble fractions obtained after sonication and centrifugation, showed that most of the recombinant protein was in the soluble fraction. Both GWD3-SBD constructs were purified to apparent homogeneity using a two step procedure including initial capture by His-tag affinity chromatography which resulted in 50–70% pure fractions based on SDS-PAGE. As a second purification step, β-CD-Sepharose affinity chromatography was used. After elution with 10 mM β-CD both GWD-SBDs were electrophoretically pure (Fig. 8.3).
The molecular mass of the protein encoded by amino acid sequence 68-184 is 14.2 kDa and the protein encoded by amino acid sequence 68-179 is 13.6 kDa, including the His tag at the C-terminus. This is in agreement with its apparent molecular mass as estimated by SDS-PAGE analysis. When GWD3-SBD was grown at higher temperature during induction an increased amount of protein was produced. Only low amounts of the recombinant protein produced at 30 °C bound to the β-CD affinity column and a higher yield was achieved for the GWD3-SBD expressed at lower temperature. Under optimised conditions the overall yield of pure GWD3-SBD from a 3 L *E. coli* culture was typically 2 mg

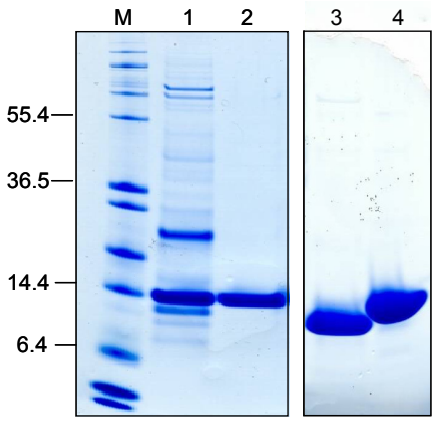


Figure 8.3; **SDS-PAGE analysis of purified *E. coli* pET28 GWD3-SBD**. Lane 1-3 GWD3-SBD (68-184); Lane 1: GWD3-SBD after His-trap purification, Lane 2: GWD3-SBD after β-CD purification. Lane 3: concentrated GWD3-SBD after β-CD purification Lane 4: concentrated GWD3-SBD (68-179) after β-CD purification. Lane M: molecular weight standard

8.1.2 GA-SBD and LAF-SBD

Construction and expression of the GA-SBD was first tested in the *E. coli* pDEST14 vector. Small scale expression test did not show clear signs of expressed protein and since the GWD3-SBD was being produced from in the pDEST17 vector instead of the pDEST14 vector, the GA-SBD was also cloned into the pDEST17 vector for expression with an N-terminal GST tag. After 5 hours of induction by 0.2% L-arabinose at 30 °C a very clear band appeared in SDS-PAGE with an estimate size of 40.8 kDa (including the GST tag). This size is in agreement with the expected size of the GA-SBD protein. Purification of the GA was achieved by affinity chromatography and the GA-SBD was highly enriched in the elution fractions and with very few contaminants. An estimate of the overall purity is in the range of 97% obtained using a single column purification strategy and the yield from 1 L culture was 2 mg GA-SBD (Appendix Fig. 13.1).

The LAF-SBD was cloned and produced from two different constructs. One representing the amino acid sequence Met^{118}-Glu^{251} and a shorter version Glu^{138}-Glu^{251}. The shorter version was selected after the construction of a homology model of the LAF-SBD, since there was no obvious fold patterns deduced from the extra sequence. After 3 hours of induction an intense band with the expected size (40.4 kDa) was obtained and verified by using antibodies against GST on a Western blot (data not shown) and the LAF-SBD was purified after a single purification step to ~95% purity (Appendix Fig. 13.1).

8.1.3 GST tag

A GST protein was produced as a control recombinant protein using the pET-41a (+) vector. The protein was purified by two steps of affinity purification. First the protein was captured on a GSTrap column and after this purification step the protein was ~85% pure. This was followed by HisTrap purification and resulted in apparent 99% pure protein (Appendix Fig. 13.2).

8.1.4 GA and GA without SBD

The GA and the GA without SBD was cloned into the pPIZα B vector and the cells were grown for 48 hours after start of induction. This resulted in a high expression level of both GA and GA without the SBD and after affinity chromatography purification they were more than 95% pure (Appendix Fig. 13.2). A typical yield from a 1 L culture was 22 mg for the GA without SBD and 18 mg for the GA.

8.1.5 GA-GWD3 fusion

The GA-GWD3 fusion was successfully cloned by overlap PCR reaction and positive transformants were selected from screening MM-starch plates. Four positive transformants were used for further small scale expression test and finally one transformant was selected based on highest expression level. A Western blot of the expression appearance profile of the recombinant protein is shown in Figure 8.4. When the sample was analysed by GA antibody two bands appeared, one with the expected size (the upper band) 71 kDa plus glycosylation and one band with a size compatible with GA without the SBD (Fig. 8.4, A). When the same expression profile was analysed with antibodies raised against the C-terminal c-myc epitope only one band appeared corresponding to the GA-GWD3 fusion full length protein (Fig. 8.4, B). This clearly demonstrates that the SBD is cleaved from the GA-GWD3 fusion during production and that the degradation product is matching with the size of the GA without the SBD. To purify the GA-GWD3 fusion a HisTrap purification method was utilised, and the fusion did bind to the column but the cleaved GWD3-SBD also bound, because it was cleaved in the N-terminal end and therefore still possesses the His tag.

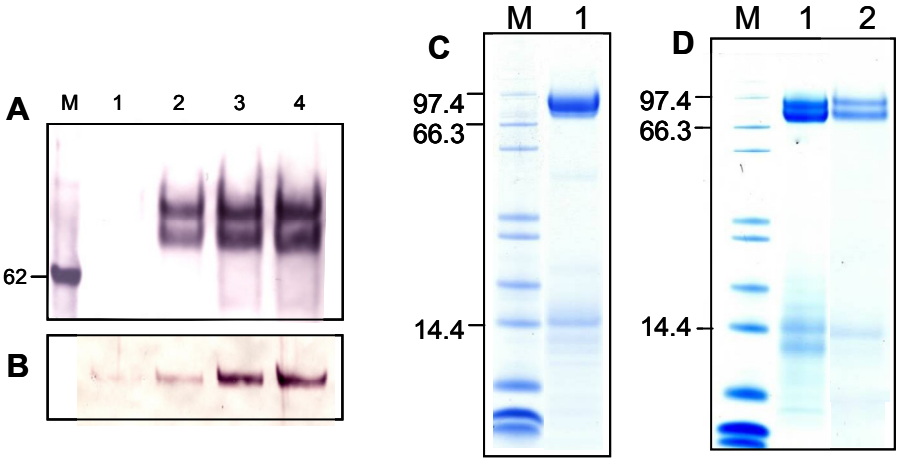


Figure 8.4: Expression and purification of the *P. pastoris* GA-GWD3 fusion.
A: Expression profile detected with GA-antibodies. Lane 1: induction start. Lane 2: 18 hours after induction. Lane 3: 22 hours after induction. Lane 4: 26 hours after induction. **B**: Same expression profile detected with c-myc antibodies. **C** and **D**: SDS-PAGE of GA-GWD3 fusion purification. Lane 1 elution from HisTrap colum. **D** Lane 1: GA-GWD3 fusion after resourceQ column purification. Lane 2: GA-GWD3 fusion after gel filtration.

Several other purification methods were used to purify the GA-GWD3 fusion from other proteins and the proteolytically cleaved GWD3-SBD. The GA-GWD3 fusion

did bind to both ResourceQ and ResourceS columns, but all eluted fractions containing the GA-GWD3 fusion also contained cleaved GA without SBD and the cleaved SBD part. A β-CD Sepharose affinity purification was tested at two different pH values (pH 5.5 and 7.5), but neither the GA-GWD3 fusion nor the degradation products of it showed any binding to the column. Several attempts were made to prevent cleavage of the construct, both during the fermentation of the culture and purification. Expression in a minimal medium (to acidify the medium and thereby minimise the activity of proteases) and in BMMY added casamino acids (adding casamino acids to the medium, has been demonstrated to lower the proteases activity on some secreted proteins (Clare *et al.,* 1991)) did not prevent cleavage and adding protease inhibitors in all buffers did not result in a lower degradation level. The fusion was also stored under different conditions; pH 5.0 and pH 7.0 and in the imidazole elution buffer after the HisTrap purification and storing the proteins at -20 degrees did not increase the amount of non-cleaved GA-GWD3 fusion protein.

8.2 Sequence motifs in CBM20

The ability of CBM20 to bind to starch seems to be associated with the presence of certain consensus residues. Originally, 11 consensus residues were suggested based on multiple alignments of 8 sequences from fungi and bacteria (Svensson *et al.,* 1989). Some of these residues are more important than others and typically, these residues are aromatic, and some are directly involved in glucan interactions. Two independent glucan binding sites are usually found in the CBM20 family. Binding site 1 contains two conserved tryptophan residues, which in the well studied *A. niger* GA-SBD are W543 and W590 (Fig. 8.2) as well as the conserved K578, E591 and N595 (Sorimachi *et al.,* 1997). Site 2 is not found in all CBM20s and has only one highly conserved amino acid residue- in GA-SBD W563, which is indirectly involved in binding. In some CBM20 sequences two tyrosines (Y527 and Y556 in GA-SBD) are assigned to binding site 2 (Sorimachi *et al.,* 1997). In addition to the original 11 consensus residues F519 in GA-SBD is 87.4% conserved in the 103 sequences analysed (Machovic *et al.,* 2005). Noticeably, this phenylalanine is neither conserved in GWD3 nor in several bacterial β-amylases (Fig. 8.2 or see review Paper IV). The deduced amino acid sequence of the GWD3-SBD shows relatively low identity to other CBM20s having 26% identity/58% similarity to *A. niger* GA-SBD. A putative GWD3-SBD from *Oryza sativa* has recently been accepted to the CBM20 family and for comparison this sequence displays 50% identity/69% similarity with the GWD3-SBD from *A. thaliana.* Tryptophans at GWD3-SBD putative binding site 1 (W102 and W142) are conserved as well as W115 (putative binding site 2) and the GA-SBD Y527 is represented by a conservative substitution as F86 in the GWD3-SBD. Y556 also found to be involved in binding site 2 in the GA-SBD is missing from the GWD3-SBD sequence that has a seven amino acid deletion. This deletion is also seen in the putative GWD3 from *Oryza sativa*. The W164 in the GWD3-SBD, is well conserved amongst the CBM20 members and mutational analysis of the W615 in GA-SBD suggested this residue plays a structural role as the W615K variant was difficult to produce (Williamson *et al.,* 1997). Three out of the eleven conserved

amino acid residues (a threonine, a leucine and a proline) are not conserved in the GWD3-SBD. The *At*GWD3 and the *Os*GWD3 has a relative high representations of Lysines (8 in the *At*GWD3 and 9 in the *Os*GWD3) compared to only 2 found in the GA-SBD. Maybe these charged amino acids could play a role in the recognition of phosphorylation in the C6 position

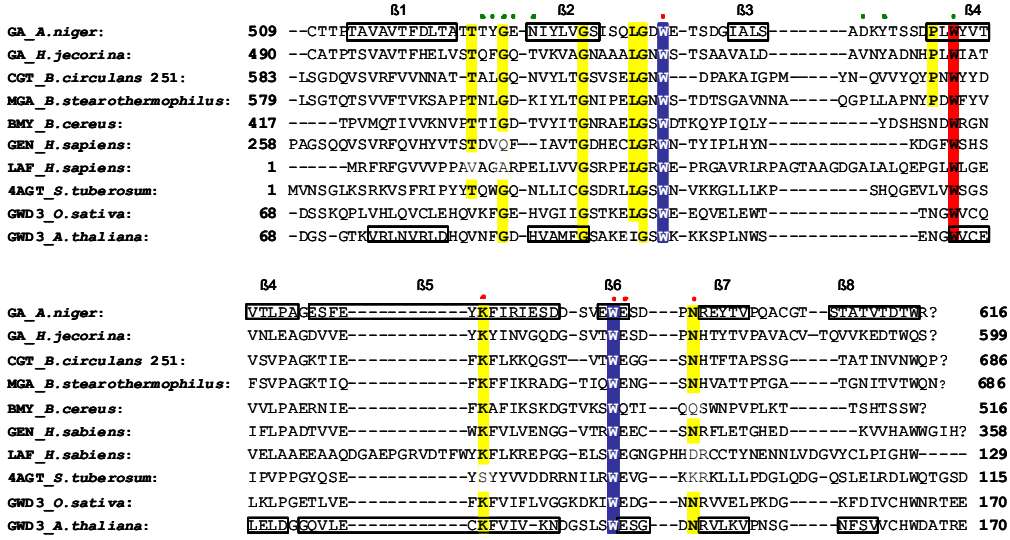


Figure 8.2 Alignment of bacterial and eukaryotic CBM20s.
Highly conserved residues in CBM20s are highlighted by colours. The highly conserved tryptophans in binding site 1 are highlighted in blue and W563 from binding site 2 in red, other conserved residues are shown in yellow. The residues involved in substrate binding are indicated by binding site 1 (red dot) and binding site 2 (green dot). The β-strands are derived from the structures of the GA-SBD and the homology model of GWD3-SBD and are shown in boxes. GA: glucoamylase, CGT: cyclodextrin glucanosyltransferase, MGA: maltogenic α-amylase, BMY: β-amylase, GEN: genethonin-1, LAF: Laforin, 4AGT: 4-α-glucanotransferase, GWD3: glucan, water dikinase 3.

The *H. sapiens* laforin sequence is even less homologous to other CBM20s than the GWD3-SBD and has only 17% identity/30% similarity with the *A. niger* GA-SBD (Fig. 8.2). The tryptophans from binding site 1 are conserved as is also the tryptophan found in binding site 2. Both tyrosines from binding site 2 in the GA-SBD are substituted by non aromatic amino acids and no aromatic amino acids are found in the sequence around binding site 2. The alignment also reveals insertions not recognised in the other CBM20 sequences and four out of the eleven conserved residues are not found in the laforin sequence.

8.3 Homology modelling

The amino acid sequence of the putative GWD3-SBD (68-170) was used in a BLAST search to identify a template for homology modelling. *A. niger* GA-SBD (PDB code: 1KUL) displayed the highest sequence identity (26% identity and 58% similarity) in the Protein Data Bank (Fig. 8.3, A). The relatively low sequence identity between target and template proteins rendered the GWD3-SBD modelling a

challenging task. However, it has been revealed that CBM20s both in eukaryotes and prokaryotes share a similar structural fold (Rodriguez-Sanoja *et al.*, 2005).
The modelled three-dimensional structure of GWD3-SBD exhibited a β-sandwich fold with an immunoglobulin-like topology (Fig. 8.3, B). In total, seven β-strands are distributed into two β-sheets. The architecture of the ligand-binding sites is structurally well conserved among the CBM20s. Both tryptophans in binding site 1 are conserved at the sequence level, and possibly at the structural level as well, as both resides are superimposable between the GA-SBD structure and the homology model of the GWD-SBD (Fig. 8.4). The assumed binding site 2 is represented by the two aromatic amino acids, F86 and W115 in the GWD3-SBD as compared with Y527 and a W563 in the GA-SBD structure, but the third aromatic residue involved in GA-SBD binding site 2 Y556 is not conserved. Binding of substrates to binding site 2 in the GA-SBD results in a conformational change were the loop between β-strand 3 and 4 shift to a position closer to the loop between β-strand 1 and 2 (Fig. 8.4). The same loop in the GWD3-SBD is very small due to the seven amino acid sequence deletion mentioned above (Fig. 8.4). The deletion in GWD3-SBD includes amino acid residue Y556, in the GA-SBD, shown to interact with β-CD, and could therefore account for differences found in binding affinities.

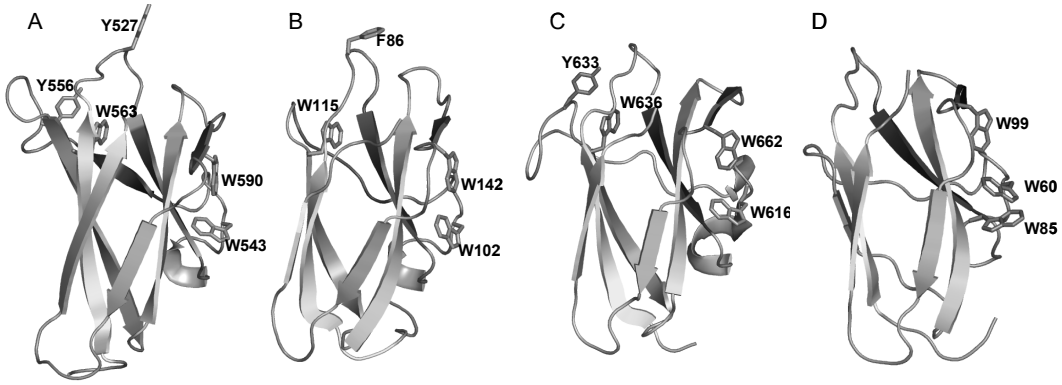


Figure 8.3. Homology models and structure of CBM20 used as template
A, *A. niger* GA-SBD (PDB code 1KUL). **B**, Homology model of *A. thaliana* GWD3-SBD. **C**, *B. circulans* strain 251 CGTase-SBD (PDB code 2DIJ). **D**, Homology model of *H. sapiens* LAF-SBD. Aromatic amino acid residues involved in the two binding sites are shown as sticks. The well conserved binding site 1 is located on the right side on each domain.

A homology model of the LAF-SBD was built on the CGTase-SBD *B. circulans* (PDB code 2DIJ). This structure displayed the highest sequence identity (19% identity and 31% similarity) to LAF-SBD in the Protein Data Bank. The overall fold of the LAF-SBD homology model also exhibited a β-sandwich fold arranged as two β-sheets of seven β-strands. Two tryptophans in binding site 1 are conserved (W60 and W85), however, a third surface exposed tryptophan (W99) is observed in close proximity of binding site 1 suggesting a differences in binding site (including three tryptophans) as compared to most other CBM20. In the sequence alignment W99 is aligned with conserved tryptophans involved in binding site 2, e.g. the GA-SBD W563 (Fig. 8.2) and this could indicate that the LAF-SBD belongs to a new clan of CBMs. When the CGTase-SBD and the LAF-SBD are aligned the root

mean square deviation (rmsd) values are 14.1 Å for only 57 aligned C_α atoms, indicating poor quality of the model due to the low homology and possible structural differences. For comparison the SBD of the glucoamylase is quite similar to the average structure of the *B. circulans* CGTase-SBD and when the two SBDs are aligned the rmsd is 1.95 Å for 91 aligned Cα atoms and they display a sequence identity (38% identity and 59% similarity).

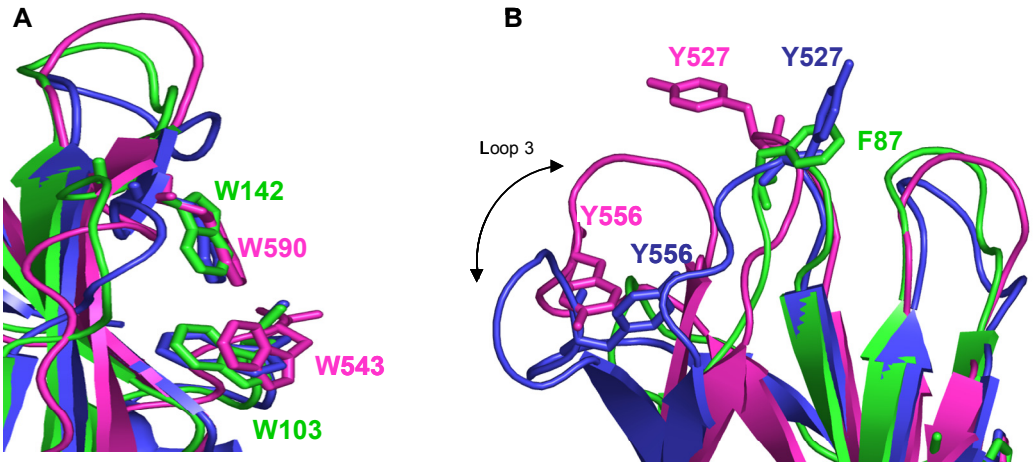


Figure 8.4 Superimposition of binding site1 and 2
A: Binding site 1 from the structure of GA-SBD in the presence of β-CD (magenta), the GA-SBD structure in its free form (blue) and the GWD3-SBD model (green). **B**: Binding site 2

8.4 Stability measurements by DSC

The conformational stability of the recombinant GWD3-$SBD_{68\text{-}184}$ produced in *E. coli* construct was evaluated by DSC and resulting in a two-step DSC profile (Fig. 5.8). The DSC curve shows that the GWD3-SBD starts to unfold at a low temperature ~20 ºC followed by a broad endothermic transition with an assigned T_m of 44.6 ºC. After this first transition, the calorimetric trace was irregular showing a broad transition between 50 and 75 ºC. The qualitative appearance and the T_m value of the first transition were reproducible within 0.6 ºC, but the second transition was not reproducible between different samples. The second observed peak was only observed in the presence of residual amounts of β-CD remaining from the purification procedure. These transitions were not visible in the GWD3-SBD samples recaptured on a HisTrap column and extensively washed on column before elution (100 CV). This suggests that β-CD binding increases the thermal stability of the domain as the re-purified samples showed roughly 5 ºC lower T_m and significantly smaller enthalpy change. All thermal transitions were completely irreversible as judged by the lack of recovery of peak area after cooling and rescanning. These results suggest that the first thermal transition represents the thermal unfolding of the GWD-SBD followed by aggregation of the unfolded protein. The small enthalpy change during the unfolding transition and the broadness of the peak are suggestive of the low conformational stability of the GWD3-SBD when present in isolated form. For example the GA-SBD from *A.*

niger has a melting temperature of 56.7 ± 0.2 ºC and is characterised by reversible unfolding – refolding (Christensen *et al.,* 1999). It is likely that the reason for the instability of the GWD-SBD is caused by the fact that the domain is expressed as an isolated domain and gets destabilised by exposure of a hydrophobic surface and lack of interactions between the GWD3-SBD and the rest of the protein. No endothermic transition could be detected for the *P. pastoris* construct indicating no stability even at low temperature (data not shown).

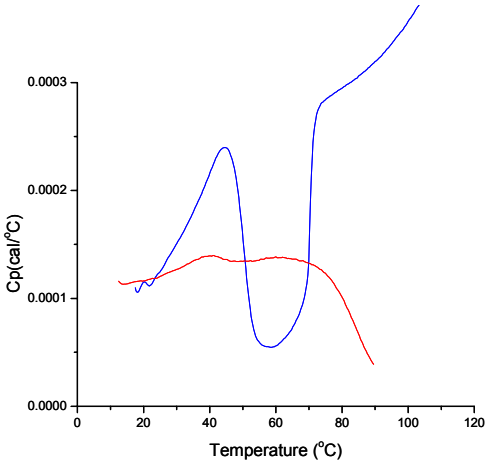


Figure 8.5: The calorimetric scan of the GWD3 SBD$_{68-184}$
Blue curve represents GWD3-SBD after dialysis. Red curve represents GWD3-SBD after recapturing and purification on a HisTrap column.

8.5 Affinity determined by SPR

Affinity towards α-, β-, and γ-CD and small maltooligosaccharides was analysed by SPR (Table 1). The SBDs were immobilized on sensor chips to levels of approximately 300 response units (RU) using the standard amine coupling method. CDs were injected over the immobilized surface at concentrations ranging from 10 µM to 10 mM (0.1 µM−1 mM in the case of the GA-SBD). To keep the GWD3-SBD protein stable for longer periods the temperature was set to 10 ºC during the SPR experiments. As compared with data collected at 25 ºC, the K_d values were very similar (data not shown) but the binding levels were maintained at high levels for several hours. The GWD3-SBD$_{68-184}$ showed strongest binding towards α-CD with a K_d of 0.22 mM at pH 5.5 and the weakest affinity towards γ-CD at pH 9.0 (Table 5.1). The affinity decreased as the pH in the buffer became more basic when α-, and γ-CD was analysed. The β-CD affinity minimum was found at neutral pH, interestingly thus at a physiologically relevant pH value. No detectable affinity for linear maltooligosaccharides (DP 3-7) or for phosphorylated β-CD was found. Different NaCl concentrations (50-500 mM) in the pH 7 buffer did not affect the binding constants. The *P. pastoris* GWD3-SBD was tested at pH 7.0 and the strongest affinity was found using α-CD and weakest affinity towards γ-CD was confirmed also by this construct. GWD3-SBD forms produced in *P. pastoris* and as a GST fusion in *E. coli* gave similar weak K_d values for β-CD at pH 7.0, as the

GWD3-SBD$_{68-184}$, excluding artefacts due to expression host or affinity tag (Table 5.1). β-CD affinity of 1.8 mM was found when the GST-GWD3-SBD fusion was analysed at pH 7. The GST tag was also tested to examine the binding contribution of the tag and unfortunately the GST tag alone also showed affinity towards β-CD with a K_d of about 8 mM. This result makes the data from the GST-GWD3-SBD complicated to analyse. The LAF-SBD was also expressed with the GST tag and analysed by SPR the domain showed a similar low binding affinity of 0.8 mM towards β-CD at pH 7.0. Data from this construct are also difficult to interpret, due to the GST tags affinity towards β-CD. A deglycosylated version of the *P. pastoris* GWD3-SBD was analysed to examine the effect of glycosylation on binding properties. The deglycosylated version had an affinity towards β-CD of 1.8 mM compared to 0.7 mM found in the glycosylated GWD3-SBD. The weaker affinity found in the deglycosylated version underlines the results found in the DSC experiments where β-CD stabilised the GWD3-SBD.

Table 5.1: Dissociation constants for GWD3-SBD and GASBD determined using SPR

CBM20 (produced in)	**Ligand**	**pH**	K_d **(mM)**
GWD3-SBD$_{68-184}$ *(E. coli)*	α-CD	5.5	0.22 ± 0.01
		7.0	0.45 ± 0.01
		9.0	0.59 ± 0.16
	β-CD	5.0	0.62 ± 0.02
		5.5	0.38 ± 0.07
		6.0	0.94 ± 0.02
		6.5	1.09 ± 0.08
		7.0	1.07 ± 0.19
		9.0	0.56 ± 0.12
	γ-CD	5.5	0.84 ± 0.66
		7.0	1.47 ± 0.40
		9.0	5.56 ± 2.07
GWD3-SBD *(P. pastoris)*	α-CD	7.0	0.4 ± 0.2
	β-CD	7.0	0.7 ± 0.2
	γ-CD	7.0	3.8 ± 2.2
GWD3-SBD *(P. pastoris)* endoH treated	β-CD	7.0	1.8 ± 0.2
GST-GWD3-SBD *(E. coli)*	β-CD	7.0	1.8*
GST-LAF-SBD *(E. coli)*	α-CD	7.0	0.8*
	β-CD	7.0	1.3*
GST-control *(E. coli)*	β-CD	7.0	8.0*
GA-SBD *(A. niger)*	β-CD	5.5	0.0075 ± 0.0001

*Data from one experiment. The rest are duplicates.

The optimum affinity for β-CD of *A. niger* GA is at pH 4.5 (Goto *et al.,* 1994) and when comparing K_d values of the GA-SBD and the GWD3-SBD at pH 5.5 a 50-fold higher affinity was found for the GA-SBD than for the GWD3-SBD (Table 5.1). The GA affinity towards β-CD has been determined by UV difference spectroscopy (K_d = 14.4 μM (Williamson *et al.,* 1997) and by NMR titration (K_d = 9 μM (Le Gal-Coëffet *et al.,* 1995) and is consistent with the present data found by SPR. Affinity differences between the two GA-SBD binding sites has been assessed for the mutants, W563K (site 2) and W590K (site 1) (Williamson *et al.,* 1997) are reflected in K_d of 28 μM and 6.4 μM, respectively. This agrees with K_{d1} of 2.0 and K_{d2} of 31.3 μM obtained by fitting a two binding site model to the GA-SBD SPR data using

Sigmaplot. In the case of the GWD3-SBD data fitting a two binding site model increased the R-values and thereby indicates a small probability that this domain also have two binding sites (Fig. 8.6).

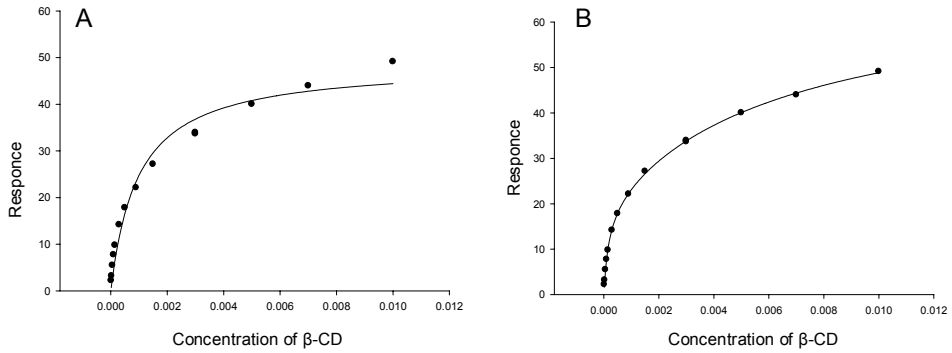


Figure 8.6: **SPR analysis of β-CD binding to *P. pastoris* GWD3-SBD.**
The difference in response units is plotted as a function of β-CD concentration. **A**: Data fitting to a 1 site model by Sigmaplot. K_d = 0.7 mM. **B**: Data fitting to a 2 site model by Sigmaplot. K_{d1} = 0.9 mM and K_{d2} = 0.043 mM

8.6 In vitro *starch binding visualised by CLSM*

Binding of fluorescent labelled GWD3-SBD to starch granules was visualised qualitatively by CLSM. A molar labelling degree of 0.2-0.5 was obtained with Fluorescein 5-EX succinimidyl ester. Adsorption of fluorescein-labelled GWD3-SBD to starch granules were demonstrated, including normal and waxy (high amylopectin content) barley and maize endosperm starch (A type crystalline polymorph) and normal potato tuber starch (B type crystalline polymorph, high phosphate). Clear fluorescence was detected on the selected starch types tested (Fig. 8.7). Distinctive binding patterns to different starches were observed, *e.g.* binding to maize and waxy barley granules (Fig. 8.7, B and C) was in specific areas on the surface, whereas binding to barley and potato starch granules was seen as a continuously distribution of fluorescence at the surface of the granules (Fig 8.7, A and E). Both GWD3- and GA-SBDs interacted internally with the maize and waxy maize, in channels and caves -characteristic of maize starches (Glaring *et al.,* 2006) and also internally in the waxy barley starch granules. More distinct surface localisation was observed for GA- than for GWD3-SBD, especially with maize starches suggesting that GWD3-SBD in that case more efficiently penetrates the granule. Adsorption of fluorescent labelled GA was investigated and clear fluorescence was seen on the surface of barley starch (Fig. 8.8, A) and after overnight incubation at 30 °C starch granules were penetrated by GA and in some case the interior granule was filled with GA (Fig. 8.8, B). Binding to maize granules

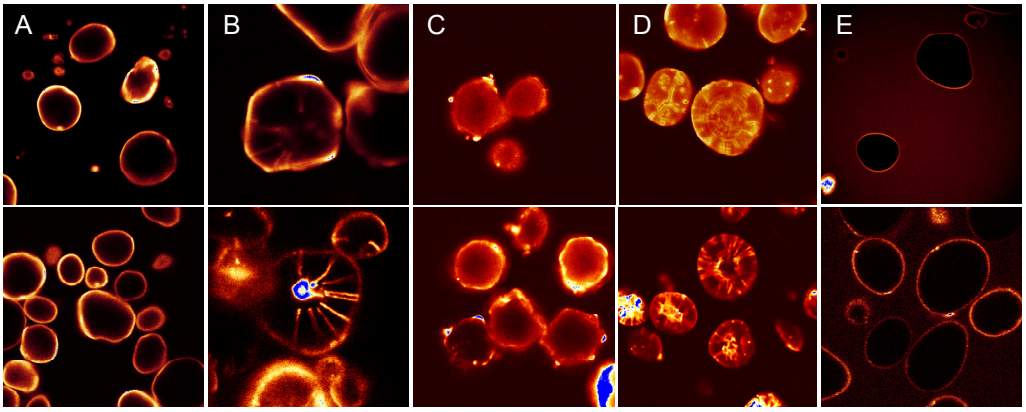


Figure 8.7: Interaction of fluorescein-labelled GWD3-SBD (top panels) and GA-SBD (bottom panels) with different starch granules.
A, barley; **B**, maize; **C**, waxy barley; **D**, waxy maize; **E**, potato.

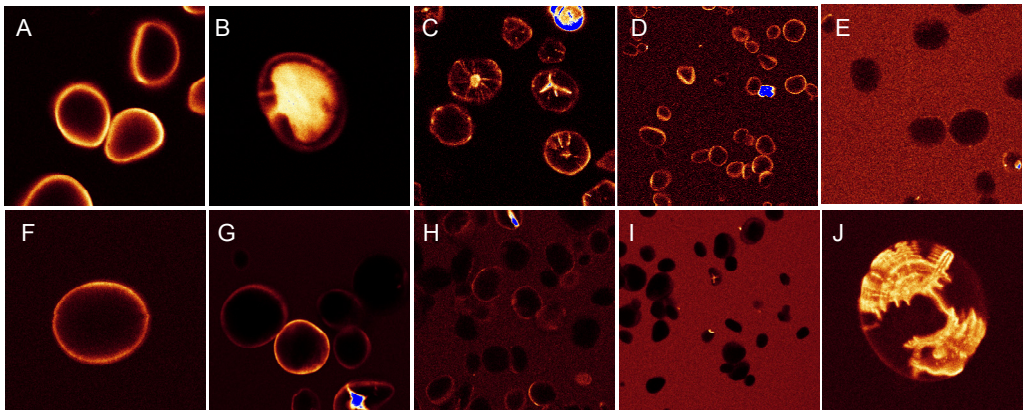


Figure 8.8: Interaction of fluorescein-labelled GA (top panels) and AMY1 (bottom panels) with different starch granules and inhibitors.
A, barley starch. **B**, barley starch after overnight incubation at 30 °C. **C** maize starch. **D** barley starch and 1 µM acarbose after overnight incubation at 30 °C. **E**, barley starch and 1 mM β-CD after overnight incubation at 30 °C. **F–I** barley starch **F**, AMY1 wt. **G**, AMY1 Y380A. **H**, AMY1 W278A/ W279A. **I**, AMY1 W278A/Y380A. **J**, AMY1 wt and maize starch after 5 hours incubation at 30 °C. (experiments A; B, D, E and F were first performed by Sauer (2001)

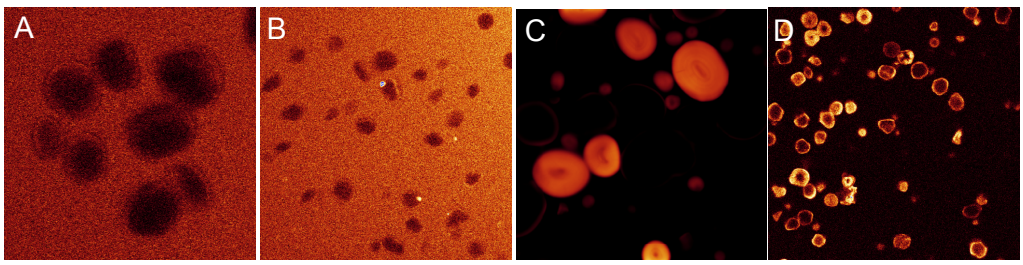


Figure 8.9: **Fluorescein labelled proteins and their interactions with raw barley starch.**
A, G2. **B**, BSA. **C**, Fluorescein free dye. **D**, GST.

was verified and the characteristic channels and cavity found in the maize starch granules was also penetrated by the GA (Fig. 8.8, C). The presence of 1 μM acarbose (a pseudotetrasaccharide inhibitor with picomolar affinity (Svensson and Sierks, 1992)) in the mixture did not affect adsorption of fluorescent labelled GA to the barley starch granules, but prevented the degradation as seen from lack of signs of damaged on the starch granules (Fig. 8.8, D). Adding 1 mM β-CD to the mixture of fluorescein labelled GA and barley starch granules resulted in inhibited binding and fluorescein labelled GA was seen in the surrounding area between starch granules and not on the starch granule surface (Fig. 8.8, E). This was expected since β-CD binds strongly to the two binding sites in the SBD with a K_d = 17.7 10^{-6} M (Goto et al., 1994) as confirmed by SPR in the present thesis project.

Barley α-amylase (AMY1) another starch degrading enzyme and AMY1 mutants at two different secondary surface binding site mutants was similarly mixed with starch granules and the adsorbtion was analysed by CLSM (Paper II). Wild type AMY1 binds to barley starch with a K_d = 0.64 ± 0.06 mg/ml (Nielsen et al., 2008) and the adsorption was verified by clear fluorescence on the granule surfaces (Fig. 8.8, F). The sugar tong mutant Y380A was able to adsorb onto the starch granules, but not with same intensity as wild type and some fluorescence was seen in the solution between the starch granules indicating poor binding (Fig. 8.8, G). Mixing the W278A/W279A mutant with barley starch granules also showed very large increase in the fraction of unbound protein and the W278A/Y380A mutant, which has mutations at both binding sites showed no signs of adsorption (Fig. 8.8 H). The reduced adsorption ability found in the secondary binding sites mutants is consistent with a starch granule adsorption analysis showing increased dissociation constants for the mutants Y380A of K_d = 5.9 ± 0.05 mg/ml, W278A/W279A, K_d = 22 ± 2 mg/ml and W278A/Y380A could not be determined (Nielsen 2006; Nielsen *et al*., 2008). Remarkably, in the close up of a maize starch granule after 5 hours incubation with AMY1 (Fig. 8.8, I) the fluorescence is distributed according to growth rings and presumably the binding capacity to the amylopectin fraction is less constrained compared to the amylose fractions and AMY1 adsorptions and degradation is initiated in the semi crystalline layers.

To verify that binding was an effect of specific interaction between the SBDs and starch granules and not only due to unspecific binding, BSA and GA without the SBD (i.e*.,* G2) were labelled with fluorescein and interaction between the proteins and starch granules was analysed by CLSM. No visible interaction could be found (Fig. 8.9, A and B) and fluorescence labelled protein was distributed in the background between the starch granules. Free dye penetrated the granules and resulted in unspecific binding (Fig. 8.9, C). The GST tag was also labelled to analyse the interaction with raw starch granules and indeed interaction was visualised (Fig. 8.9, D) and confirmed the SPR data that indicated weak affinity towards β-CD for GST (K_d = 8 mM).

8.7 Binding of GWD3-SBD to starch granules in vivo

To visualise *in vivo* binding of the GWD3-SBD to starch granules an YFP-GWD3-SBD fusion was prepared. The GWD3-SBD contained the region encoding residues (1-201), including the chloroplast transit peptide and the SBD. The expression in tobacco of the gene encoding the fusion protein was driven by the 35S constitutive promoter. The YFP-GWD3-SBD fusion was analysed by CLSM and transient expression was visualised in tobacco leaves (Fig. 8.10). This verified the affinity for starch granules of the GWD3-SBD *in vivo* since YFP fluorescence of the transient expression product was localised in the chloroplasts (red chlorophyll autofluorescence) and was essentially confined to the starch granules.

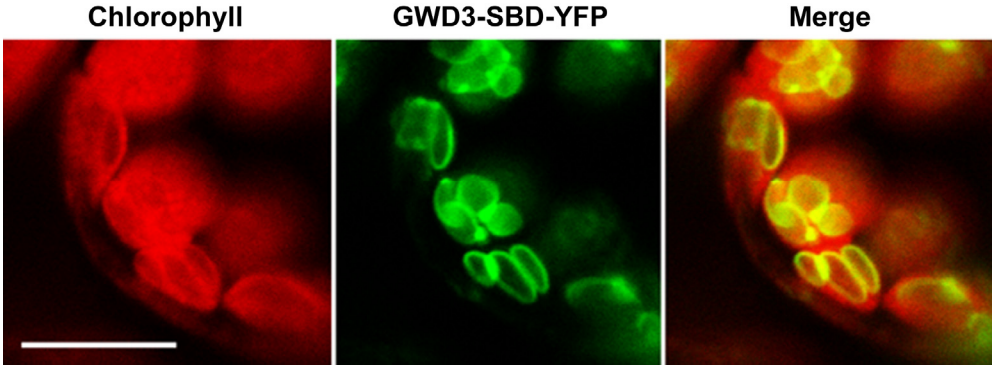


Figure 8.10: GWD3-SBD binding *in vivo.*
YFP-GWD3-SBD transiently expressed with its *A. thaliana* GWD3 transit peptide in tobacco leaf mesophyll cells, showing binding at the surface of starch granules. Scale bar: 10 µm.

9 Discussion

The present data summarises the effort that has been done in this thesis to produce the functional GWD3-SBD in isolated form to characterise its biochemical function, and to produce this SBD as a fusion with the *A. niger* glucoamylase. This turned out to be a very challenging task due to stability problems that will be discussed below. Several constructs were made using different tags and different expression hosts, and the result was a functional and stable enough form of GWD3-SBD to generate reliable binding data. The results obtained highlighted the status of the GWD3 CBM20 as a novel type with different properties than other characterised CBM20s (Paper III).

Analysis of CBM annotation and choice of constructs

Initially, the polypeptide spanning amino acid residues 68-170 was chosen based on previously defined domain annotations (Baunsgaard *et al*., 2005). Another publication that came out simultaneously defined a shorter length of the CBM spanning residues 84-155 (Kötting et al., 2005). The assignment of the chloroplastic transit peptide is predicted with the cleavage site between amino acid residue 52 and 53 (Fig. 7.1). The polypeptide stretch separating the predicted signal peptide from the annotated N-terminus of the SBD resembles linker sequences often rich in serines and threonines and terminating with four consecutive positively charged residues (64-67). It is unclear what the role of this sequence is, but it is unlikely to be part of the CBM according to sequence alignments. This, however, does not exclude the possibility that this part may play a role in anchoring the domain to the full-length *At*GWD3. The annotation of the C-terminal end of the SBD was strictly based on sequence alignments, where the R168 in the GWD3 aligns with the C-terminal R616 of the homologous *A. niger* CBM20. Therefore all constructs made at the start of the project for expression in *E. coli* and *P. pastoris* were based on this annotation. Other factors that complicate the annotation of the GWD3-SBD borders are the lack of structural characterisation of its catalytic domain precluding prediction of inter-domain contacts, and the fact that the majority of characterised CBM20s are C-terminally located, whereas the GWD3 is N-terminal (Paper IV).

All these initial constructs encoding isolated modules as well as fusions with either a GST tag or with the GA from *A. niger* showed poor stability together with purification problems. This led to the conclusion that the annotated domain borders did not yield a stable isolated module. Analysis of the amino acid residues in the sequence at the C-terminal end suggested that the C-terminal end of GWD3-SBD may be longer than deduced from the sequence alignment. This hypothesis is supported by the lack of amino acids like glycine or proline typically present in secondary structure transitions (Fig. 7.1). This motivated the design of two additional constructs which span the amino acid sequences 68-179 and 68-184 and contain possible secondary structure changes. No further support to the annotation, however, was obtained by secondary structure predictions which gave different results depending on the servers used. The various forms of GWD3-SBD and their production are discussed further below.

***E. coli* gateway constructs**

The *E. coli* gateway expression system was chosen because it gives the possibility to easily shuffle constructs between different expression vectors encoding different tags. Firstly the expression was tested using a pDEST14 vector, which does not encode any tags. From that system no visible signs of expression were detected in the form of an over-expressed protein band shown in SDS-PAGE. No antibodies were available for GWD3-SBD, thus detection by Western blotting was not possible. The GWD3-SBD sequence was next by recombination transferred to a pDEST15 vector, which encodes an N-terminal GST tag. pDEST15 was chosen for several reasons; i) the GST tag is three times bigger than GWD3-SBD and may mimic the catalytic domain of GWD3 and thereby stabilise the SBD which may be unstable when expressed just as the isolated domain, ii) the GST tag displays catalytic activity simplifying quantification in starch binding assays; and iii) commercial antibodies are available against GST. Recombinant protein was obtained and an optimised purification resulted in 70–80% pure GWD3-SBD. In addition to the GWD3-SBD-GST fusion protein, another protein corresponding in size to the GST tag alone and presumably some GWD3-SBD degradation products were co-purified. Several purification strategies summarised in the results part were tested but without improved purity. During the period of purification optimisation, binding site mutants were made using the pDONR201 vector or the pDEST15 and these mutants (W142T and F86A) were over-expressed as judged by total cell protein SDS-PAGE analysis, but no solubility tests were performed.

To overcome difficulties with unwanted cleavage of the GST tag and formation of other degradation products, the construct was transferred to the pDEST17 vector encoding an N-terminal His tag. This allows another purification strategy to be examined and it also makes it possible to use the GWD3-SBD binding site mutants made in the compatible pDONR201. Soluble expression of pDEST17 GWD3-SBD was achieved, but the His tag was cleaved off during the production and this GWD3-SBD form was not successfully purified and moreover it was not stable in solution during storage. None of these constructs resulted in a protein able to bind to β-CD affinity resins, which indicated that the expressed proteins were not correctly folded and/or were not displaying binding activity perhaps due to oligomerisation or due to the presence of the GST tag. In parallel, expression of two other CBM20 domains, the *A. niger* GA-SBD and the *H. sapiens* LAF-SBD, were also done in the pDEST15 vector for comparison and GST with no domain fusion was made to perform control experiments. Expression and purification of the GA- and the LAF-SBDs were successful, but because a functional GWD3 SBD construct was the main objective further characterisation of these other SBD domains was not pursued.

***P. pastoris* constructs**

One of the objectives of the project was to exchange the GA-SBD in the full length fungal GA with the GWD3-SBD. Expression in the methylotrophic yeast *P. pastoris* is suitable for functional catalytic GA domain (Fierobe et al., 1997). This host is able to form disulfide bridges and carry out glycosylations both known to be present in GA. For this reason, a *P. pastoris* construct of the isolated GWD3-SBD

was also made with the same domain borders (68-170) as the above described GWD3-SBD and joined to a C-terminal His-tag.
The fusion between the *A. niger* GA catalytic domain and the GWD3-SBD$_{68-170}$ were the focus of the planned study, but the full length GA, a truncated GA form lacking the CBM20, and the isolated GA-CBM20 were made as controls. The expressed GWD3-SBD unfortunately underwent proteolytic degradation possibly at the N-terminal part since the GA degradation products co-purified on the HisTrap column as verified by Western blotting using antibodies against the C-terminal c-myc epitope. It cannot be excluded that degradation happens after the elution from the HisTrap column. It was not possible to prevent this degradation by any of the tested purification methods accounted for in the results part, or the use of protease inhibitors and degradation continued during storage.
The high level of recombinant GWD3-SBD secreted from *P. pastoris* made it possible to produce enough partially purified GWD3-SBD to evaluate its structural integrity by DSC. This analysis confirmed that the form exhibited poor structural stability as thermograms obtained in this case were very similar to baseline scans and no clear thermal transition was observed. It cannot be ruled out that the protein lost its stability due to degradation between purification and performing the DSC analysis, still it was a clear indication that this form was not useful for further studies.
It was not surprising that similar problems with degradation and instability were observed in the GA-GWD3 fusion construct. Induction at 22 °C as compared to 30 °C resulted in higher yields, but the SBD was still cleaved off and GWD3-SBD degradation products and full length fusion protein were not separated by HisTrap purification due to the C-terminal His-tag and other tested purification strategies failed to yield pure fusion protein. Neither the GWD3-SBD nor the fusion produced in *P. pastoris* bound to the β-CD affinity resin, which also could be due to the folding or stability problems. The remaining control constructs lacking the GWD3-SBD were expressed successfully and purified to homogeneity. These results made it very clear that the work with the GA-GWD3 fusion was not meaningful since the isolated domain was not stable enough and hence work at this stage was focused on trying to make a better construct of the GWD3-SBD as an isolated domain.

***E. coli* constructs encoding a longer CBM border sequence at the C-terminus**
A breakthrough in the present project came with the design of two new *E. coli* constructs based on analysis of the primary structure as discussed above. This analysis suggested that the GWD3-SBD may have a longer C-terminal region as judged by the presence of hydrophobic residues downstream of the formerly annotated sequence and lack of an apparent secondary structure shift in the previously annotated sequence with C-terminus at amino acid residue 170. The two new constructs encoding amino acid sequence 68-179 and 68-184 were expressed in *E. coli* and possessed a C-terminal His tag. Both GWD3-SBD forms were purified to apparent homogeneity using a two step procedure including initial capture by His-tag affinity chromatography on His-Trap, followed by β-CD-Sepharose affinity chromatography. After elution with 10 mM β-CD both GWD-SBDs were electrophoretically pure. The GWD3-SBD$_{68-184}$ was found by DSC to be the most

stable form and was therefore used in the subsequent characterisation (Paper III). The purification by β-CD strongly indicated that the domain was correctly folded and had binding activity. Another indirect confirmation of the ligand binding came from the DSC analysis after β-CD affinity purification where the GWD3-SBD seemed to be very importantly stabilised by the β-CD, which was apparently retained in the 3.5 kDa molecular weight cut off dialysis bags (Molecular weight of β-CD= 1135)
This form of recombinant GWD3-SBD, however, still had lower stability that the isolated GA-SBD (see paper III). The structural fold within a CBM20 family is conserved and the average length of CBMs varies only within a narrow range. This small variation is likely to be within loop regions and cannot be very large if the structural fold has to be preserved. Based on this, it is most likely that the borders of the longest form are correct. The low stability of this SBD form is not due to folding problems but might reflect the absence of stabilising interactions between the SBD and the rest of the GWD3 enzyme, when it is expressed as the isolated domain. Examples of CBM20s that are intimately packed on their catalytic domains are seen in CGTase and the β-amylase, and CBM20s that are connected to the catalytic domain via long polypeptide linkers are seen in the GA (Paper IV).

Characterisation of the most stable GWD3-SBD
It has been suggested that the GWD3 acts preferentially on amylopectin already phosphorylated at the C-6 position of glucosyl residues. Therefore it is likely that the substrate for GWD3-SBD is transient starch granules in the leaves during the biosynthesis, which are possibly phosphorylated by the action of GWD1 (Baunsgaard et al., 2005; Kötting et al., 2005). The lack of characterised plant CBM20 members and the low homology to other CBM20s make bioinformatics based analysis of the properties of the GWD3-SBD not straightforward. An analysis of pI values of isolated SBDs and full length enzymes did not show a significant difference in the case of the GWD3 or comparison with other enzymes containing CBM20s. Analysis of the local charge distribution, however, and comparing to the homologue GA-SBD, the GWD3-SBD have a significantly higher number of positively charged residues surrounding binding site 1. This may reflect preference for the negatively charged phosphorylated glycosyl residues in α-glucan chains. The low affinity of the GWD3-SBD did not allow any discrimination between the different starches in the binding assays. The GWD3-SBD did not show any binding to phosphorylated β-CD, which has on average four glucosyl residues phosphorylation, thus a much higher degree of modification than in the natural starch, having a degree of phosphorylation is substantially lower (0.1-1% of the glucose units in amylopectin contains covalently bound phosphate).
The two tryptophans found in the binding site 1 in most CBM20 are conserved in the GWD3-SBD sequence (W102 and W142) and it is suggested that F86 and W115 are in involved in binding site 2. The most conspicuous difference is the 7 amino acid deletion in the GWD3-SBD compared to the GA-SBD sequence (amino acid residue 553-558) in the flexible loop of site 2 containing the binding aromatic residue Y556 (*A. niger* GA numbering) and importantly implicated in binding site 2 (see paper III). The homology model confirmed the conservation of binding site 1

and supported the hypothesis that amino acid residue deletions in the flexible loop could affect the GWD3-SBD binding mechanisms or affinity.
Oligosaccharide binding studies performed by SPR confirmed affinity of GWD3-SBD towards β-CD as reflected also from the affinity purification. The GWD3-SBD_{68-184} was analysed in more detail since it was concluded to be correctly folded and functional based on the above arguments, generally K_d values, however were in the range of 0.4–2 mM for all the GWD3 forms when β-CD was used as ligand. Also affinity towards α- and γ-CD was demonstrated as well as a small effect of the pH value, when β-CD was used as ligand, with the lowest affinity being observed at neutral pH.
The reason why the *P. pastoris* GWD3-SBD and the *E.coli* gateway GWD3-SBDs failed to bind to the β-CD resin could be either too poor stability or aggregation occurring during the purification procedure.
The GA-SBDs affinity towards β-CD at pH 5.5 was 50-fold higher than of GWD3-SBD, thus clearly demonstrating the difference in capacity of the two domains in interacting with small soluble oligosaccharides. The difference was also evident in the affinity towards starch granules as it was not possible to assay starch granule adsorption for the GWD3-SBD. However, the affinity towards starch granules of GWD3-SBD was unquestionably, albeit qualitatively, demonstrated both *in vitro* and *in vivo* using confocal laser scanning microscopy (paper III) on GWD3-SBD mixtures with suspended starch granule and expression in tobacco, respectively. It is not evident why this clear binding to starch granules cannot be verified by classical binding assays, but it may be the low binding capacity or preferential binding to a low abundance type of binding sites at the starch granule which causes this behaviour. According to this hypothesis, if the GWD3-SBD has a preference for low abundance sites which are few relative to other sites used by classical CBM20 domains, flurophore-labelled GWD3-SBD bound at the low abundance sites would still contribute fluorescence at the granule surface as compared to the bulk solvent.
The physiological relevance of an SBD with low affinity or binding capacity demonstrated for the plant CBM20 is substantiated by another plant SBD of CBM45 found in GWD1 (Mikkelsen *et al.,* 2006) and CBM53 in starch synthase III (Valdez *et al.,* 2008), both having low affinity for starch grnaules. As for GWD3 these enzymes are predominantly found in soluble form when extracted. Furthermore, GWD1 binding to starch *in planta* is stronger in the dark (Ritte *et al.,* 2000b; Mikkelsen *et al.,* 2005). By analogy GWD3-SBD can confer reversible and dynamic binding of GWD3 to starch granules in response to physiological needs.

10 Conclusions

Initially the present project was based on the hypothesis that the *At*GWD3-SBD was a typical CBM20 that could replace the fungal SBD in the starch degrading GA from *A. niger*. The aim then was to evaluate the effect of the plant SBD on enzymatic performance of the GA. The plan turned out to be very challenging experimentally due to the instability in the hosts *E. coli* and *P. pastoris* of the *At*GWD3-SBD form designed based on the original annotation of the domain region in the GWD3. The *P. pastoris* produced recombinant GWD3-SBD was stabilised to some extent presumably by glycosylation, but was sensitive to proteolytic activity and the resulting GWD3-SBD degradation products were difficult to separate from the intact GWD3-SBD. More problematically the purification of, these various recombinant GWD3-SBDs on β-cyclodextrin–Sepharose was not possible probably due to their low stability in solution. Two new forms were designed based on sequence analysis, which were both C-terminally extended compared to the original forms. Both these new forms were purified to homogeneity using a two step protocol. Importanly, β-cyclodextrin affinity purification used in the second step confirmed that SBD to have affinity for the ligand and thus indicated that the recombinant GWD3-SBD form was correctly folded and functionally active. *E. coli* GWD3-$SBD_{68\text{-}184}$ was found to be the most stable form by differential scanning calorimetry analysis, although it still had a rather poor conformational stability. Furthermore, this calorimetric analysis showed that the domain was significantly stabilised by β-cyclodextrin.

The reason for the low thermal stability of the GWD3-SBD as compared to the *A. niger* GA-SBD could be that the domain interacts tightly with the neighbouring domains in the full-length GWD3 enzyme resulting in loss of conformational integrity when produced isolated from the rest of the protein. Thus no polypeptide linker seems to connect the SBD with the catalytic domain of GWD3 in contrast to the situation found for *A. niger* GA of which the produced SBD had much superior stability.

Despite the low stability of the GWD3-$SBD_{68\text{-}184,}$ this form allowed the characterisation of the binding affinity towards soluble ligands such as .cyclodextrin using SPR. The GWD3-$SBD_{68\text{-}184}$ had a K_d of ~1 mM towards β-CD when tested at physiological relevant pH 6-7, which is described as a low affinity binding. The low affinity presented here was surprising, since other characterised CBM20s are described as moderate affinity binders with K_d values in the low micromolar range.

Affinity towards starch granules was visualised *in vitro* and *in vivo* by using CLSM. Adsorption assays to starch granules were, however, unsuccessful confirming the low affinity and suggesting a very low binding capacity. Further research is required to establish the structural features responsible for this low affinity.

Sequence analysis and homology modelling have shown that the flexible loop shown to be important for binding at the high affinity binding site 2 in CBM20 of *A. niger* GA is 7 residues shorter in the GWD3-SBD. This deletion compared to GA-SBD includes the binding site residue Y556. These features may explain the low binding capacity found in the GWD3-SBD. Recent literature suggests that other SBDs present in regulatory plant enzymes such as the *potato* glucan, water dikinase

1 CBM45 and the *Arabidopsis* starch synthase III CBM53 also display low affinity towards starch granules. This suggests that the low affinity of these SBDs may be important in mediating dynamic and regulated binding to starch required for the function of the connected catalytic domains. In other words this strengthens the hypothesis that this is a new group of CBMs, most likely belonging to a group of regulated intracellular enzymes, having completely different role than SBDs from microbial extracellular enzymes, e.g. α-amylases, CGTases and glucoamylases.

10.1 Concluding remarks

In recent years, increased insight in the diurnal starch metabolism revealed a regulatory cascade with starch phosphorylation being a key event. In spite the importance of the GWD3 several aspects of its mechanism remain unknown. The size of the full length enzyme, the presence of several domains and possible flexibility usually present in regulatory enzymes render more detailed work and structural characterisation of the full length enzyme. On the other hand, working with the isolated SBD also presented limitations in understanding the precise role of the SBD in the context of the full length enzyme. The work presented in this thesis, however, clearly confirms the starch binding activity of the SBD and strongly suggests that this SBD and possible other similar domains might have evolved a different binding mechanism to fit the regulatory role of the enzymes they are connected to. In order to understand the role of this SBD in more detail mutations *in vivo* in the full length enzyme and subsequent analysis of the effect of this on starch metabolism *in planta* is recommended.

11 Reference List

Ao,Z. and Jane,J. (2007). Characterization and modeling of the A- and B-granule starches of wheat, triticale, and barley. Carbohydr. Polym. *67*, 46-55.

Ball,S., Guan,H.P., James,M., Myers,A., Keeling,P., Mouille,G., Buleon,A., Colonna,P., and Preiss,J. (1996). From glycogen to amylopectin: a model for the biogenesis of the plant starch granule. Cell *86*, 349-352.

Ball,S.G., van de Wal,M.H.B.J., and Visser,R.G.F. (1998). Progress in understanding the biosynthesis of amylose. Trends Plant Sci. *3*, 462-467.

Ball,S.G. and Morell,M.K. (2003). From bacterial glycogen to starch: understanding the biogenesis of the plant starch granule. Annu. Rev. Plant Biol. *54*, 207-233.

Ballicora,M.A., Iglesias,A.A., and Preiss,J. (2004). ADP-glucose pyrophosphorylase: A regulatory enzyme for plant starch synthesis. Photosynth. Res. *79*, 1-24.

Baunsgaard,L., Lütken,H., Mikkelsen,R., Glaring,M.A., Pham,T.T., and Blennow,A. (2005). A novel isoform of glucan, water dikinase phosphorylates pre-phosphorylated α-glucans and is involved in starch degradation in *Arabidopsis*. Plant J. *41*, 595-605.

Bertoft,E. (2004). On the nature of categories of chains in amylopectin and their connection to the super helix model. Carbohydr. Polym. *57*, 211-224.

Bertoft,E. (2007a). Composition of building blocks in clusters from potato amylopectin. Carbohydr. Polym. *70*, 123-136.

Bertoft,E. (2007b). Composition of clusters and their arrangement in potato amylopectin. Carbohydr. Polym. *68*, 433-446.

Bird,A.R., Brown,I.L., and Topping,D.L. (2000). Starches, resistant starches, the gut microflora and human health. Curr. Issues Intest. Microbiol. *1*, 25-37.

Blennow,A., Bay-Smidt,A.M., Wischmann,B., Olsen,C.E., and Møller,B.L. (1998). The degree of starch phosphorylation is related to the chain length distribution of the neutral and the phosphorylated chains of amylopectin. Carbohydr. Res. *307*, 45-54.

Blennow,A., Bay-Smidt,A.M., Olsen,C.E., and Møller,B.L. (2000). The distribution of covalently bound phosphate in the starch granule in relation to starch crystallinity. Int. J. Biol. Macromol. *27*, 211-218.

Blennow,A., Bay-Smidt,A.M., and Bauer,R. (2001). Amylopectin aggregation as a function of starch phosphate content studied by size exclusion chromatography and on-line refractive index and light scattering. Int. J. Biol. Macromol. *28*, 409-420.

Blennow,A., Engelsen,S.B., Nielsen,T.H., Baunsgaard,L., and Mikkelsen,R. (2002). Starch phosphorylation: a new front line in starch research. Trends Plant Sci. *7*, 445-450.

Blennow,A., Hansen,M., Schulz,A., Jørgensen,K., Donald,A.M., and Sanderson,J. (2003). The molecular deposition of transgenically modified starch in the starch granule as imaged by functional microscopy. J. Struct. Biol. *143*, 229-241.

Buléon,A., Colonna,P., Planchot,V., and Ball,S. (1998). Starch granules: structure and biosynthesis. Int. J. Biol. Macromol. *23*, 85-112.

Caspar,T., Lin,T.P., Kakefuda,G., Benbow,L., Preiss,J., and Somerville,C. (1991). Mutants of *Arabidopsis* with altered regulation of starch degradation. Plant Physiol *95*, 1181-1188.

Chia,T., Thorneycroft,D., Chapple,A., Messerli,G., Chen,J., Zeeman,S.C., Smith,S.M., and Smith,A.M. (2004). A cytosolic glucosyltransferase is required for conversion of starch to sucrose in *Arabidopsis* leaves at night. Plant J. *37*, 853-863.

Christensen,T., Svensson,B., and Sigurskjold,B.W. (1999). Thermodynamics of reversible and irreversible unfolding and domain interactions of glucoamylase from *Aspergillus niger* studied by differential scanning and isothermal titration calorimetry. Biochemistry *38*, 6300-6310.

Clare,J.J., Romanos,M.A., Rayment,F.B., Rowedder,J.E., Smith,M.A., Payne,M.M., Sreekrishna,K., and Henwood,C.A. (1991). Production of mouse epidermal growth factor in yeast: high-level secretion using *Pichia pastoris* strains containing multiple gene copies. Gene *105*, 205-212.

Colleoni,C., Dauvillée,D., Mouille,G., Buléon A., Gallant,D., Bouchet,B., Morell,M., Samuel,M., Delrue,B., d'Hulst,C., Bliard,C., Nuzillard,J.M., and Ball,S. (1999a). Genetic and biochemical evidence for the involvement of α-1,4 glucanotransferases in amylopectin synthesis. Plant Physiol *120*, 993-1004.

Colleoni,C., Dauvillée,D., Mouille,G., Morell,M., Samuel,M., Slomiany,M.C., Liénard,L., Wattebled,F., d'Hulst,C., and Ball,S. (1999b). Biochemical characterization of the *chlamydomonas reinhardtii* α-1,4 glucanotransferase supports a direct function in amylopectin biosynthesis. Plant Physiol *120*, 1005-1014.

Coppin,A., Varre,J.S., Lienard,L., Dauvillee,D., Guerardel,Y., Soyer-Gobillard,M.O., Buleon,A., Ball,S., and Tomavo,S. (2005). Evolution of plant-like crystalline storage polysaccharide in the protozoan parasite *Toxoplasma gondii* argues for a red alga ancestry. J. Mol. Evol. *60*, 257-267.

Critchley,J.H., Zeeman,S.C., Takaha,T., Smith,A.M., and Smith,S.M. (2001). A critical role for disproportionating enzyme in starch breakdown is revealed by a knock-out mutation in *Arabidopsis*. Plant J. *26*, 89-100.

Delatte,T., Umhang,M., Trevisan,M., Eicke,S., Thorneycroft,D., Smith,S.M., and Zeeman,S.C. (2006). Evidence for distinct mechanisms of starch granule breakdown in plants. J. Biol. Chem. *281*, 12050-12059.

Edwards,A., Fulton,D.C., Hylton,C.M., Jobling,S.A., Gidley,M., Rossner,U., Martin,C., and Smith,A.M. (1999). A combined reduction in activity of starch synthases II and III of potato has novel effects on the starch of tubers. Plant J. *17*, 251-261.

Engelsen,S.B., Madsen,A.O., Blennow,A., Motawia,M.S., Møller,B.L., and Larsen,S. (2003). The phosphorylation site in double helical amylopectin as investigated by a combined approach using chemical synthesis, crystallography and molecular modeling. FEBS Lett. *541*, 137-144.

Evers,A.D. (1971). Scanning electron microscopy of wheat starch. III. Granule development in the endosperm. Starch/Stärke *23*, 157-162.

Fierobe,H.P., Mirgorodskaya,E., Frandsen,T.P., Roepstorff,P., and Svensson,B. (1997). Overexpression and characterization of *Aspergillus awamori* wild-type and mutant glucoamylase secreted by the methylotrophic yeast *Pichia pastoris*: comparison with wild-type recombinant glucoamylase produced using *Saccharomyces cerevisiae* and *Aspergillus niger* as hosts. Protein Expr. Purif. *9*, 159-170.

Fincher,G.B. (1989). Molecular and cellular biology associated with endosperm mobilization in germinating cereal grains. Annu. Rev. Plant Physiol. Plant Mol. Biol. *40*, 305-346.

French,D. (1973). Chemical and physical properties of starch. J. Anim Sci. *37*, 1048-1061.

Gallant,D.J., Bouchet,B., and Baldwin,P.M. (1997). Microscopy of starch: evidence of a new level of granule organization. Carbohydr. Polym. *32*, 177-191.

Geera,B.P., Nelson,J.E., Souza,E., and Huber,K.C. (2006). Granule bound starch synthase I (GBSSI) gene effects related to soft wheat flour/starch characteristics and properties. Cereal Chem. *83*, 544-550.

Glaring,M.A., Koch,C.B., and Blennow,A. (2006). Genotype-specific spatial distribution of starch molecules in the starch granule: a combined CLSM and SEM approach. Biomacromolecules *7*, 2310-2320.

Glaring,M.A., Zygadlo,A., Thorneycroft,D., Schulz,A., Smith,S.M., Blennow,A., and Baunsgaard,L. (2007). An extra-plastidial α-glucan, water dikinase from *Arabidopsis* phosphorylates amylopectin in vitro and is not necessary for transient starch degradation. J. Exp. Bot. *58*, 3949-3960.

Goto,M., Tanigawa,K., Kanlyakrit,W., and Hayashida,S. (1994). The mechanism of binding of glucoamylase 1 from *Aspergillus-awamori* var. *kawachi* to cyclodextrins and raw starch. Biosci. Biotechnol. Biochem. *58*, 49-54.

Gubler,F., Raventos,D., Keys,M., Watts,R., Mundy,J., and Jacobsen,J.V. (1999). Target genes and regulatory domains of the GAMYB transcriptional activator in cereal aleurone. Plant J. *17*, 1-9.

Hanashiro,I., Abe,J., and Hizukuri,S. (1996). A periodic distribution of the chain length of amylopectin as revealed by high-performance anion-exchange chromatography. Carbohydr. Res. *283*, 151-159.

Hanashiro,I., Tagawa,M., Shibahara,S., Iwata,K., and Takeda,Y. (2002). Examination of molar-based distribution of A, B and C chains of amylopectin by fluorescent labeling with 2-aminopyridine. Carbohydr. Res. *337*, 1211-1215.

Henrissat,B. and Davies,G. (1997). Structural and sequence-based classification of glycoside hydrolases. Curr. Opin. Struct. Biol. *7*, 637-644.

Hizukuri,S. (1985). Relationship between the distribution of the chain length of amylopectin and the crystalline structure of starch granules. Carbohydr. Res. *141*, 295-306.

Hizukuri,S. (1986). Polymodal distribution of the chain lengths of amylopectins, and its significance. Carbohydr. Res. *147*, 342-347.

Huber,K.C. and BeMiller,J.N. (2000). Channels of maize and sorghum starch granules. Carbohydr. Polym. *41*, 269-276.

Hussain,H., Mant,A., Seale,R., Zeeman,S., Hinchliffe,E., Edwards,A., Hylton,C., Bornemann,S., Smith,A.M., Martin,C., and Bustos,R. (2003). Three isoforms of isoamylase contribute different catalytic properties for the debranching of potato glucans. Plant Cell *15*, 133-149.

Imberty,A. (1988). A revisit to the three-dimensional structure of B-type starch. Biopolymers *27*, 1205-1221.

Imberty,A., Chanzy,H., Perez,S., Buleon,A., and Tran,V. (1988). The double-helical nature of the crystalline part of A-starch. J. Mol. Biol. *201*, 365-378.

Jane,J., Kasemsuwan,T., Leas,S., Zobel,H., and Robyt,J.F. (1994). Anthology of starch granule morphology by scanning electron microscopy. Starch/Stärke *46*, 121-129.

Janecek,S. (2002). A motif of a microbial starch-binding domain found in human genethonin. Bioinformatics *18*, 1534-1537.

Jenkins,P.J., Cameron,R.E., and Donald,A.M. (1993). A universal feature in the structure of starch granules from different botanical sources. Starch/Stärke *45*, 417-420.

Jobling,S. (2004). Improving starch for food and industrial applications. Curr. Opin. Plant Biol. *7*, 210-218.

Jobling,S.A., Westcott,R.J., Tayal,A., Jeffcoat,R., and Schwall,G.P. (2002). Production of a freeze-thaw-stable potato starch by antisense inhibition of three starch synthase genes. Nat. Biotechnol. *20*, 295-299.

Kalé,K., Skeel,R., Bhandarkar,M., Brunner,R., Gursoy,A., Krawetz,N., Phillips,J., Shinozaki,A., Varadarajan,K., and Schulten,K. (1999). NAMD2: Greater scalability for parallel molecular dynamics. J. Comput. Phys. *151*, 283-312.

Kammerer,B., Fischer,K., Hilpert,B., Schubert,S., Gutensohn,M., Weber,A., and Flugge,U.I. (1998). Molecular characterization of a carbon transporter in plastids from heterotrophic tissues: the glucose 6-phosphate/phosphate antiporter. Plant Cell *10*, 105-117.

Kaplan,F. and Guy,C.L. (2005). RNA interference of *Arabidopsis* β-amylase 8 prevents maltose accumulation upon cold shock and increases sensitivity of PSII photochemical efficiency to freezing stress. Plant J. *44*, 730-743.

Kötting,O., Pusch,K., Tiessen,A., Geigenberger,P., Steup,M., and Ritte,G. (2005). Identification of a novel enzyme required for starch metabolism in Arabidopsis leaves. The phosphoglucan, water dikinase. Plant Physiol. *137*, 242-252.

Kozlov,S.S., Blennow,A., Krivandin,A.V., and Yuryev,V.P. (2007). Structural and thermodynamic properties of starches extracted from GBSS and GWD suppressed potato lines. Int. J. Biol. Macromol. *40*, 449-460.

Kuipers,A., Jacobsen,E., and Visser,R. (1994). Formation and deposition of amylose in the potato tuber starch granule are affected by the reduction of granule-bound starch synthase gene expression. Plant Cell *6*, 43-52.

Le Gal-Coëffet,M.F., Jacks,A.J., Sorimachi,K., Williamson,M.P., Williamson,G., and Archer,D.B. (1995). Expression in *Aspergillus niger* of the starch-binding domain of glucoamylase. Comparison with the proteolytically produced starch-binding domain. Eur. J. Biochem. *233*, 561-567.

Lloyd,J.R., Blennow,A., Burhenne,K., and Kossmann,J. (2004). Repression of a novel isoform of disproportionating enzyme (stDPE2) in potato leads to inhibition of starch degradation in leaves but not tubers stored at low temperature. Plant Physiol *134*, 1347-1354.

Lloyd,J.R., Kossmann,J., and Ritte,G. (2005). Leaf starch degradation comes out of the shadows. Trends Plant Sci. *10*, 130-137.

Lorberth,R., Ritte,G., Willmitzer,L., and Kossmann,J. (1998). Inhibition of a starch-granule-bound protein leads to modified starch and repression of cold sweetening. Nat. Biotechnol. *16*, 473-477.

Machovic,M. and Janecek,S. (2006). The evolution of putative starch-binding domains. FEBS Lett. *580*, 6349-6356.

Machovic,M., Svensson,B., MacGregor,E.A., and Janecek,S. (2005). A new clan of CBM families based on bioinformatics of starch-binding domains from families CBM20 and CBM21. FEBS J. *272*, 5497-5513.

Manners,D.J. (1989). Recent developments in our understanding of amylopectin structure. Carbohydr. Polym. *11*, 87-112.

Manners,D.J. (1991). Recent developments in our understanding of glycogen structure. Carbohydr. Polym. *16*, 37-82.

Marshall,J., Sidebottom,C., Debet,M., Martin,C., Smith,A.M., and Edwards,A. (1996). Identification of the major starch synthase in the soluble fraction of potato tubers. Plant Cell *8*, 1121-1135.

Mikkelsen,R., Baunsgaard,L., and Blennow,A. (2004). Functional characterization of α-glucan,water dikinase, the starch phosphorylating enzyme. Biochem. J. *377*, 525-532.

Mikkelsen,R. and Blennow,A. (2005). Functional domain organization of the potato α-glucan, water dikinase (GWD): evidence for separate site catalysis as revealed by limited proteolysis and deletion mutants. Biochem. J. *385*, 355-361.

Mikkelsen,R., Mutenda,K.E., Mant,A., Schurmann,P., and Blennow,A. (2005). α-glucan, water dikinase (GWD): a plastidic enzyme with redox-regulated and coordinated catalytic activity and binding affinity. Proc. Natl. Acad. Sci. U. S. A *102*, 1785-1790.

Mikkelsen,R., Suszkiewicz,K., and Blennow,A. (2006). A novel type carbohydrate-binding module identified in α-glucan, water dikinases is specific for regulated plastidial starch metabolism. Biochemistry *45*, 4674-4682.

Minassian,B.A., Ianzano,L., Meloche,M., Andermann,E., Rouleau,G.A., Gado-Escueta,A.V., and Scherer,S.W. (2000). Mutation spectrum and predicted function of laforin in Lafora's progressive myoclonus epilepsy. Neurology *55*, 341-346.

Mouille,G., Maddelein,M.L., Libessart,N., Talaga,P., Decq,A., Delrue,B., and Ball,S. (1996). Preamylopectin processing: a mandatory step for starch biosynthesis in plants. Plant Cell *8*, 1353-1366.

Myers,A.M., Morell,M.K., James,M.G., and Ball,S.G. (2000). Recent progress toward understanding biosynthesis of the amylopectin crystal. Plant Physiol. *122*, 989-997.

Nabeshima,E.H. and Grossmann,V.E. (2001). Functional properties of pregelatinized and cross-linked cassava starch obtained by extrusion with sodium trimetaphosphate. Carbohydr. Polym. *45*, 347-353.

Nielsen,T.H., Wischmann,B., Enevoldsen,K., and Møller,B.L. (1994). Starch phosphorylation in potato tubers proceeds concurrently with *de novo* biosynthesis of starch. Plant Physiol *105*, 111-117.

Nielsen,M.M. (2006) Digestion and interaction of starches with α-Amylases: I: Mutational analysis of carbohydrate binding sites in barley α-Amylases 1. II: In vitro starch digestion of legumes. Ph.D. Thesis, Enzyme and protein chemistry, BioCentrum-DTU.

Nielsen,M.M., Seo,E.S., Bozonnet,S., Aghajari,N., Robert,X., Haser,R., and Svensson,B. (2008). Multi-site substrate binding and interplay in barley α-amylase 1. FEBS Lett. *582*, 2567-2571.

Niittyla,T., Messerli,G., Trevisan,M., Chen,J., Smith,A.M., and Zeeman,S.C. (2004). A previously unknown maltose transporter essential for starch degradation in leaves. Science *303*, 87-89.

Nour-Eldin,H.H., Hansen,B.G., Norholm,M.H.H., Jensen,J.K., and Halkier,B.A. (2006). Advancing uracil-excision based cloning towards an ideal technique for cloning PCR fragments. Nucleic Acids Research *34*.

Oostergetel,G.T. and van Bruggen,E.F.J. (1993). The crystalline domains in potato starch granules are arranged in a helical fashion. Carbohydr. Polym. *21*, 7-12.

Parker,R. and Ring,S.G. (2001). Aspects of the physical chemistry of starch. J. Cereal Sci. *34*, 1-17.

Plotnikov,V.V., Brandts,J.M., Lin,L.N., and Brandts,J.F. (1997). A new ultrasensitive scanning calorimeter. Anal. Biochem. *250*, 237-244.

Ral,J.P., Derelle,E., Ferraz,C., Wattebled,F., Farinas,B., Corellou,F., Buléon,A., Slomianny,M.C., Delvalle,D., d'Hulst,C., Rombauts,S., Moreau,H., and Ball,S. (2004). Starch division and partitioning. A mechanism for granule propagation and maintenance in the picophytoplanktonic green alga *Ostreococcus tauri*. Plant Physiol *136*, 3333-3340.

Ramon,M., Rolland,F., Thevelein,J.M., Van,D.P., and Leyman,B. (2007). ABI4 mediates the effects of exogenous trehalose on *Arabidopsis* growth and starch breakdown. Plant Mol. Biol. *63*, 195-206.

Reimann,R., Ritte,G., Steup,M., and Appenroth,K.J. (2002). Association of α-amylase and the R1 protein with starch granules precedes the initiation of net starch degradation in turions of *Spirodela polyrhiza*. Physiol. Plant. *114*, 2-12.

Reimann,R., Hippler,M., Machelett,B., and Appenroth,K.J. (2004). Light induces phosphorylation of glucan water dikinase, which precedes starch degradation in turions of the duckweed *Spirodela polyrhiza*. Plant Physiol. *135*, 121-128.

Reimann,R., Ziegler,P., and Appenroth,K.J. (2007). The binding of α-amylase to starch plays a decisive role in the initiation of storage starch degradation in turions of *Spirodela polyrhiza*. Physiol. Plant *129*, 334-341.

Ritte,G., Eckermann N, Haebel S, Lorberth R, and Steup,M. (2000a). Compartmentation of the starch-related R1 protein in higher plants. Starch/Stärke *52*, 179-185.

Ritte,G., Lorberth,R., and Steup,M. (2000b). Reversible binding of the starch-related R1 protein to the surface of transitory starch granules. Plant J. *21*, 387-391.

Ritte,G., Lloyd,J.R., Eckermann,N., Rottmann,A., Kossmann,J., and Steup,M. (2002). The starch-related R1 protein is an α-glucan, water dikinase. Proc. Natl. Acad. Sci. U. S. A *99*, 7166-7171.

Ritte,G., Scharf,A., Eckermann,N., Haebel,S., and Steup,M. (2004). Phosphorylation of transitory starch is increased during degradation. Plant Physiol. *135*, 2068-2077.

Ritte,G., Heydenreich,M., Mahlow,S., Haebel,S., Kotting,O., and Steup,M. (2006). Phosphorylation of C6- and C3-positions of glucosyl residues in starch is catalysed by distinct dikinases. FEBS Lett. *580*, 4872-4876.

Roldan,I., Wattebled,F., Mercedes,L.M., Delvalle,D., Planchot,V., Jimenez,S., Perez,R., Ball,S., d'Hulst,C., and Merida,A. (2007). The phenotype of soluble starch synthase IV defective mutants of *Arabidopsis thaliana* suggests a novel function of elongation enzymes in the control of starch granule formation. Plant J. *49*, 492-504.

Sauer, J. (2001). Stability and function of interdomain linker variants and a fluorescence resonance energy transfer system set-up for the two-domain enzyme glucoamylase 1 from *Aspergillus niger*. Ph.D. Thesis, Carlsberg Laboratory, Department of chemistry, Copenhagen.

Sali,A. and Blundell,T.L. (1993). Comparative protein modelling by satisfaction of spatial restraints. J. Mol. Biol. *234*, 779-815.

Scheidig,A., Frohlich,A., Schulze,S., Lloyd,J.R., and Kossmann,J. (2002). Downregulation of a chloroplast-targeted β-amylase leads to a starch-excess phenotype in leaves. Plant J. *30*, 581-591.

Schwall,G.P., Safford,R., Westcott,R.J., Jeffcoat,R., Tayal,A., Shi,Y.C., Gidley,M.J., and Jobling,S.A. (2000). Production of very-high-amylose potato starch by inhibition of SBE A and B. Nat. Biotechnol. *18*, 551-554.

Shure,M., Wessler,S., and Fedoroff,N. (1983). Molecular identification and isolation of the Waxy locus in maize. Cell *35*, 225-233.

Smith,A.M., Denyer,K., and Martin,C.R. (1995). What controls the amount and structure of starch in storage organs? Plant Physiol *107*, 673-677.

Smith,A.M. (2001). The biosynthesis of starch granules. Biomacromolecules *2*, 335-341.

Sorimachi,K., Jacks,A.J., Le Gal-Coëffet,M.F., Williamson,G., Archer,D.B., and Williamson,M.P. (1996). Solution structure of the granular starch binding domain of glucoamylase from *Aspergillus niger* by nuclear magnetic resonance spectroscopy. J. Mol. Biol. *259*, 970-987.

Sorimachi,K., Le Gal-Coëffet,M.F., Williamson,G., Archer,D.B., and Williamson,M.P. (1997). Solution structure of the granular starch binding domain of *Aspergillus niger* glucoamylase bound to β-cyclodextrin. Structure *5*, 647-661.

Steichen,J.M., Petty,R.V., and Sharkey,T.D. (2008). Domain characterization of a 4-α-glucanotransferase essential for maltose metabolism in photosynthetic leaves. J. Biol. Chem. *283*, 20797-20804.

Steup,M. and Schächtele,C. (1981). Mode of glucan degradation by purified phosphorylase forms from spinach leaves. Planta *153*, 351-361.

Svensson,B., Jespersen,H., Sierks,M.R., and MacGregor,E.A. (1989). Sequence homology between putative raw-starch binding domains from different starch-degrading enzymes. Biochem. J. *264*, 309-311.

Svensson,B. and Sierks,M.R. (1992). Roles of the aromatic side chains in the binding of substrates, inhibitors, and cyclomalto-oligosaccharides to the glucoamylase from *Aspergillus niger* probed by perturbation difference spectroscopy, chemical modification, and mutagenesis. Carbohydr. Res. *227*, 29-44.

Tabata,S. and Hizukuri,S. (1971). Studies on starch phosphate, Part 2. Isolation of glucose-3-phosphate and maltose phosphate by acid hydrolysis of potato starch. Starch/Stärke *23*, 267-272.

Takeda,Y. and Hizukuri,S. (1981). Re-examination of the action of sweet-potato β-amylase on phosphorylated (1-4)-α- D-glucan. Carbohydr. Res. *89*, 174-178.

Takeda,Y. and Hizukuri,S. (1987). Structures of branched molecules of amyloses of various origins, and molar fractions of branched and unbranched molecules. Carbohydr. Res. *165*, 139-145.

Tatge,H., Marshall,J., Martin,C., Edwards,E.A., and Smith,A.M. (1999). Evidence that amylose synthesis occurs within the matrix of the starch granule in potato tubers. Plant. Cell Environ. *22*, 543-550.

Valdez,H.A., Busi,M.V., Wayllace,N.Z., Parisi,G., Ugalde,R.A., and Gomez-Casati,D.F. (2008). Role of the N-terminal starch-binding domains in the kinetic properties of starch synthase III from *Arabidopsis thaliana*. Biochemistry *47*, 3026-3032.

Vasanthan,T. and Hoover,R. (1992). A comparative study of the composition of lipids associated with starch granules from various botanical sources. Food Chemistry *43*, 19-27.

Viksø-Nielsen,A., Blennow,A., Jørgensen,K., Kristensen,K.H., Jensen,A., and Møller,B.L. (2001). Structural, physicochemical, and pasting properties of starches from potato plants with repressed r1-gene. Biomacromolecules *2*, 836-843.

Voinnet,O., Rivas,S., Mestre,P., and Baulcombe,D. (2003). An enhanced transient expression system in plants based on suppression of gene silencing by the p19 protein of tomato bushy stunt virus. Plant J. *33*, 949-956.

Waigh,T.A., Perry,P., Riekel,C., Gidley,M.J., and Donald,A.M. (1998). Chiral side-chain liquid-crystalline polymeric properties of starch. Macromolecules *31*, 7980-7984.

Wattebled,F., Dong,Y., Dumez,S., Delvalle,D., Planchot,V., Berbezy,P., Vyas,D., Colonna,P., Chatterjee,M., Ball,S., and d'Hulst,C. (2005). Mutants of Arabidopsis lacking a chloroplastic isoamylase accumulate phytoglycogen and an abnormal form of amylopectin. Plant Physiol *138*, 184-195.

Weise,S.E., Weber,A.P., and Sharkey,T.D. (2004). Maltose is the major form of carbon exported from the chloroplast at night. Planta *218*, 474-482.

Wiesenborn,D.P., Orr,P.H., Caspar,H.H., and Tacke,B.K. (1994). Potato starch paste behavior as related to some physical/chemical properties. J. Food. Sci. *59*, 644-648.

Williamson,M.P., Le Gal-Coëffet,M.F., Sorimachi,K., Furniss,C.S., Archer,D.B., and Williamson,G. (1997). Function of conserved tryptophans in the *Aspergillus niger* glucoamylase 1 starch binding domain. Biochemistry *36*, 7535-7539.

Wilson,J.D., Bechtel,D.B., Todd,T.C., and Seib,P.A. (2006). Measurement of wheat starch granule size distribution using image analysis and laser diffraction technology. Cereal Chemistry *83*, 259-268.

Yu,T.S., Kofler,H., Häusler,R.E., Hille,D., Flugge,U.I., Flugge,U.I., Smith,A.M., Kossmann,J., Lloyd,J., Ritte,G., Steup,M., Lue,W.L., Chen,J.C., and Weber,A. (2001). The Arabidopsis *sex1* mutant is defective in the R1 protein, a general regulator of starch degradation in plants, and not in the chloroplast hexose transporter. Plant Cell *13*, 1907-1918.

Yu,T.S., Zeeman,S.C., Thorneycroft,D., Fulton,D.C., Dunstan,H., Lue,W.L., Hegemann,B., Tung,S.Y., Umemoto,T., Chapple,A., Tsai,D.L., Wang,S.M., Smith,A.M., Chen,J., and Smith,S.M. (2005). α-Amylase is not required for breakdown of transitory starch in Arabidopsis leaves. J. Biol. Chem. *280*, 9773-9779.

Yu,X., Houtman,C., and Atalla,R.H. (1996). The complex of amylose and iodine. Carbohydr. Res. *292*, 129-141.

Zeeman,S.C., Umemoto,T., Lue,W.L., Au-Yeung,P., Martin,C., Smith,A.M., and Chen,J. (1998). A mutant of Arabidopsis lacking a chloroplastic isoamylase accumulates both starch and phytoglycogen. Plant Cell *10*, 1699-1712.

Zeeman,S.C., Thorneycroft,D., Schupp,N., Chapple,A., Weck,M., Dunstan,H., Haldimann,P., Bechtold,N., Smith,A.M., and Smith,S.M. (2004). Plastidial α-glucan phosphorylase is not required for starch degradation in Arabidopsis leaves but has a role in the tolerance of abiotic stress. Plant Physiol *135*, 849-858.

Zhang,X., Myers,A.M., and James,M.G. (2005). Mutations affecting starch synthase III in Arabidopsis alter leaf starch structure and increase the rate of starch synthesis. Plant Physiol *138*, 663-674.

Zheng,G.H. and Sosulski,F.H. (1998). Determination of water separation from cooked starch and flour pastes after refrigeration and freeze-thaw. J. Food. Sci. *63*, 134-139.

PAPER I

Starch

Recent Progress in Biopolymer and Enzyme Technology

EDITORS:
Piotr TOMASIK
Eric BERTOFT
Andreas BLENNOW

COVERING DESIGN:
Wojciech Ciesielski

Balicka Street 122
30 149 Kraków
tel. (48-12) 662 47 54

ISBN 978-83-902699-7-X

Printed in Poland

Starch

Recent Progress in Biopolymer and Enzyme Technology

Piotr Tomasik
Eric Bertoft
Andreas Blennow
(Editors)

POLSKIE TOWARZYSTWO
TECHNOLOGÓW ŻYWNOŚCI
ODDZIAŁ MAŁOPOLSKI

30-149 Kraków, ul. Balicka 122

Contents

This publication was fully sponsored by the Polish Ministry of Science and High Education

OUR SPONSORS:

1) Wawel Spółka Akcyjna, Kraków

2) Bahlsen Sp. z o.o., Skawina

3) Zakłady Przemysłu Cukierniczego Skawa Spółka Akcyjna, Wadowice

4) Merck Sp. zo.o., Warszawa

5) Ramboll Polska Sp. z o.o.

6) Grupa Maspex Wadowice

Preface

This year, as every even year since 2001 Kraków hosted again the International Starch Convention Kraków – Moscow. This year, 2008, the ISC XVI was organized by Agricultural University of Cracow, Polish Food Technologists' Society – Małopolska Branch, Russian Academy of Sciences, Institute of Biochemical Physics, Division of Storage and Processing of Farming Stuff of Russian Academy of Agricultural Sciences, All-Russian Institute of Starch Products, and Association of Russian Starch and Glucose Manufacturers.

This international scientific event attracted almost 100 participants representing 13 nations of the World. During the 4-day conference the participants contributed with 20 plenary lectures and 55 oral/poster presentations covering many facets of starch including structural problems of starch and related polysaccharides, their biological origin and production, their physical, chemical and enzymatic modifications and technological processes useful in polysaccharide technology. In this book, we have selected contributions we think is of special interest to readers with interest in starch research and technology. We hope that the topics presented in this issue will help to inspire scientists, engineers and other actors in the starch community to better understanding of starch and to further develop starch-based solutions for the future. We hope to see your participation in the following XVII International Starch Convention scheduled for 2009 in Moscow.

PIOTR TOMASIK

ERIC BERTOFT

ANDREAS BLENNOW

Chapter 7

Exploring New Plant Carbohydrate Binding Modules (CBMs) from Glucan, Water Dikinase

Camilla Christiansen[1,2], Maher Abou Hachem[2], Esben Friis[3], Martin J. Baumann[2], Mikkel A. Glaring[1], Anders Viksø-Nielsen[3], Bent W, Sigurskjold[4,] Birte Svensson[2] and Andreas Blennow[1]

[1]*Department of Plant Biology and Biotechnology, Faculty of Life Sciences, University of Copenhagen, DK-1871 Frederiksberg, Denmark, e-mail:abl@life.ku.dk*

[2]*Enzyme and Protein Chemistry, Department of Systems Biology, Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark*

[3]*Novozymes A/S, DK-2880 Bagsvaerd, Denmark*

[4]*Department of Biology, University of Copenhagen, DK-2100 Copenhagen Ø, Denmark*

ABSTRACT

Starch binding domains (SBDs) are distinct starch-recognizing modules frequently present in starch active enzymes. SBDs are assigned in carbohydrate binding module (CBM) families based on primary structure similarities and conserved structural folds. SBDs from CBM family 20 are encountered in archaea, bacteria, and eukaryota and occur together with a variety of catalytic domains in α-amylases, β-amylases, cyclodextrin glucosyl transferases, glucoamylases, pullulanases, glucan, water dikinases (GWDs), and other enzymes. Typically, SBDs facilitate binding of extracellular hydrolytic enzymes to starch granules and they are considered as important motifs for developing enzymes with improved and new activities for efficient degradation of starch. Members of the recently discovered family CBM45 are exclusively present in

plastidial α-amylases, the role of which is still unknown and in GWDs implicated in starch phosphorylation. Interestingly, the plastidial GWD from *Arabidopsis thaliana* GWD3 has an SBD belonging to CBM20. The GWD3 SBD was cloned and produced for further characterisation and comparison of binding properties with the well characterized CBM20 of glucoamylase (GA) from *Aspergillus niger*. Binding to the low molecular weight starch mimic β-cyclodextrin and other soluble ligands was measured using surface plasmon resonance analysis, and confocal laser scanning microscopy was used to visualise binding to starch granules. The *Arabidopsis thaliana* GWD3-SBD displayed 70 fold lower affinity to β-cyclodextrin than the GA-SBD. The low affinity of these intracellular plant CBMs as compared to most other starch binding domains including the *A. niger* CBM20 suggests reversible or regulated binding of the plant SBD-containing enzymes. This is the first report of an isolated plant CBM20 member and it highlights the presence of functional differences within CBM20.

Key words: starch binding domain, glucan water dikinase, homology modelling, surface plasmon resonance, differential scanning calorimetry, confocal laser scanning microscopy

INTRODUCTION

Starch binding enzymes. Adsorption of enzymes to starch granules is a key event in the action of starch metabolic enzymes. During starch biosynthesis and degradation in the plant, enzymes display reversible binding to starch granules as directed by e.g. photosynthesis. One recent example is the diurnal partitioning of the starch phosphorylator glucan, water dikinase 1 (GWD1) between the stroma of the plastid and the starch granule surface [1]. This mechanism provides regulation of starch metabolism during degradation and biosynthesis. Binding is effectuated by recognition sites on the enzyme surface, which may be an integral part of the catalytic domain, as for example in the granule bound starch synthase or the starch branching enzymes, involved in starch biosynthesis [2,3]. Other enzymes, however, contain domains dedicated to binding onto granular starch. These starch binding domains (SBDs) are classified in different carbohydrate binding module (CBM) families; CBM20, 21, 25, 26, 34, 41, 45 and 48 [4]; (see http://www.cazy.org/fam/acc_CBM.html). Extracellular starch hydrolases secreted by e.g. fungi or bacteria bind strongly to starch granules to ensure efficient degradation [5].

In the cyclodextrin glycosyltransferase (CGTase) from *Bacillus circulans* the SBD constitutes an integral module in the multi-domain enzyme [6]. Other members of these CBM families are separated from the catalytic domain by a linker sequence that can be *O*-glycosylated as in the case of glucoamylase (GA) from *Aspergillus niger* [7].

SBD function. It has been suggested that the function of SBDs is to attach onto granular starch and thereby maintaining the enzyme in close proximity to the substrate and increasing local enzyme concentration at the substrate surface [8]. This leads to an enhanced enzyme activity and a more rapid degradation of the polysaccharide. However, the precise impact of SBDs on starch degradation remains to be elucidated. Without an SBD, starch hydrolytic enzymes show low activity on starch granules, whereas the activity on soluble substrates, e.g. amylopectin and amylose, is not affected substantially [9,10]. In a proposed mechanism [11] the SBDs can "unwind" α-glucan helical conformations on the starch granule surface resulting in a higher hydrolytic rate [9]. Besides their affinity for starch granules, SBDs can bind maltoheptaose- and cyclodextrins [12,13] SBDs retain their affinity for starch, even when separated from the rest of the protein [12-14] and seen from a biotechnological point of view SBDs present interesting perspectives. Expression of SBDs as single modules in potato tubers thus generates structurally modified starch [15], and enzymes fused to an SBD achieve improved and more efficient degradation properties [10,16].

SBD structure. SBDs are located at the N- or C-termini of the enzyme and compared to catalytic domains, CBM20 modules are relatively small being between 90-130 residues long [17,18]. The SBDs ability to bind to starch seems to be associated with the presence of some consensus residues (Fig. 1). Not all of the consensus residues proposed originally [17] must be present for achieving starch binding, but the two tryptophans forming the substrate binding site 1 seems to be of essential importance [12].

```
                                T  G      G    LG W                     P  W
1KUL   509  -CTTPTAVAVTFDLTATTTYGENIYLVGSISQLGDWET--SDGIALSADKYTS-SDPLWYV
2VN4   490  PCATPTSVAVTFHELVSTQFGQTVKVAGNAAALGNWST--SAAVALDAVNYAD-NHPLWIA
1CDG   583  LSGDQVSVRFVVNNAT-TALGQNVYLTGSVSELGNWDP--AKAIGPMYNQVVY-QYPNWYY
1CGT   580  LTGDQVTVRFVVNNAS-TTLGQNLYLTGNVAELGNWSTG-STAIGPAFNQVIH-QYPTWYY
1B90   417  ----TPVMQTIVVKNVPTTIGDTVYITGNRAELGSWDTK-QYPIQLYYDSHSN----DWRG
2ZOB     6  ----SGPSQVAFEIRGTLLPGEVFAICGSCDALGNWNP--QNAVALLPENDTG-ESMLWKA
1PAM   582  -TGDQVTVRFVINNAT-TALGQNVFLTGNVSELGNWDP--NNAIGPMYNQVVY-QYPTWYY
1QHO   579  LSGTQTSVVFTVKSAPPTNLGDKIYLTGNIPELGNWSTDTSGAVNNAQGPLLAPNYPDWFY
GWD3    68  -DGSGTKVRLNVRLDHQVNFGDHVAMFGSAKEIGSWKK--KSPLNWSEN--------GWVC

                     K                         W    N
1KUL   TVTLPAGESFEYKFIR-----------IESDDSVEWESDPNREYTVPQACGTSTATVTDTWR□   616
2VN4   TVNLEAGDVVEYKYINV----------GQDGSVTWESDPNHTYTVPAVACVTQVVKEDTWQS□  599
1CDG   DVSVPAGKTIEFKFLKK----------QGSTVTWEGGSNHTFTAPSSG---TATINVNWQP□  686
1CGT   DVSVPAGKQLEFKFFKK----------NGSTITWEGGSNHTFTAPSSG---TATINVNWQP□  684
1B90   NVVLPAERNIEFKAFIK----------SKDGTVKSWQTIQQSWNPVPLKT----TSHTSSW□    516
2ZOB   TIVLSRGVSVQYRYFKGYFLEPKTIGGPCQVIVHKWETHLQPRSITPLESEIIIDDGQFGI     120
1PAM   DVSVPAGQTIEFKFLKK----------QGSTVTWEGGANRTFTTPTSGTATVNVNWQP□     686
1QHO   VFSVPAGKTIQFKFFIK----------RADGTIQWENGSNHVATTPTGATGNITVTWQN□     686
GWD3   ELELDGGQVLECKFVIV----------KNDGSLSWESGDNRVLKVPNSG---NFSVVCHWDATR 168
```

Figure 1. Amino acid sequence alignment of selected representatives of CBM20.

The alignment includes CBM20 sequences with solved 3D structures and the GWD3-SBD sequence. The 11 consensus residues highly conserved in CBM20s are shown above the alignment. The two tryptophans of the starch binding site 1 are highlighted in grey. Squares signify the C-terminal residues of the proteins. For abbreviations of protein sources their PDB code are used execpt for the GWD3-SBD. 1KUL, Aspergillus niger glucoamylase; 2VN4, Hypocrea jecorina glucoamylase; 1CDG, Bacillus circulans 251 cyclodextrin glucan transferase; 1CGT, Bacillus circulans 8 cyclodextrin glucan transferase; 1B90, Bacillus cereus var. mycoides β-amylase; 2ZB90, Homo sapiens FLJ11085; 1PAM, Bacillus sp. 1011 cyclodextrin glucantransferase;1QHO, Geobacillus stearothermophilus Maltotetraose-forming amylase; GWD3, Arabidopsis thaliana glucan water dikinase 3. The sequences were aligned with the program ClustalW.

While amino acid sequences are not well conserved among different SBDs, the overall protein fold is conserved and crystallographic and NMR studies so far reveal that SBDs are rather rigid structures [6,19]. The SBDs form an open-sided *β*-barrel structure with an overall topology showing seven to eight anti-parallel *β*-strands arranged into two major *β*-sheets (Fig. 2) [19,20]. Typically, aromatic amino acids like tryptophan and tyrosine in conserved positions are involved in carbohydrate binding [14,21]. SBDs often contain two separate sugar-binding sites (Fig. 2) providing bivalent binding. Each site has two or three accessible aromatic amino acid side chains interacting with sugar residues in enzyme/ligand complexes [19]. Site 1 is shallower and more surface exposed compared to site 2, while site 2 has two or three tyrosine residues situated on a flexible loop of the SBD. The site 2 undergoes a significant structural change upon binding, permitting the SBD to interact with starch in various orientations [12].

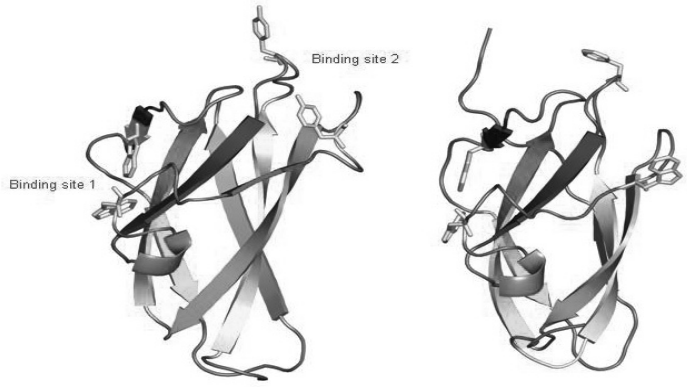


Figure 2. The NMR structure of GA-SBD (PDB code 1KUL) left) used as template for the homology model of the *Arabidopsis thaliana* GWD3-SBD (right). The aromatic amino acids involved in the two carbohydrate binding sites are shown in stick representation [19].

The intracellular SBDs. Most of the characterised SBDs are encountered in extracellular starch metabolising enzymes providing efficient attack on starch granules and have strong affinity for starch [5]. However, certain enzymes involved in intracellular starch metabolism in plants also possess SBDs including a newly discovered family, CBM45 [22]. This study reported a rather weak affinity (K_d = 1.2 mg/mL) towards soluble starch determined using affinity electrophoresis [22]. Currently CBM45 are found only as intracellular modules on enzymes from the plant kingdom. CBM45 occurs as an N-terminal tandem domain in plastidial amylases, GWD1 and GWD2 (Fig. 3). Currently less than 40 CBM45 sequences are known. SBDs from the CBM20 family display a vast phylogenetic diversity as these modules are found in archaea, bacteria, fungi, plants and mammals [23]. Several members of this family are well characterised and display high affinity towards insoluble starch [21,24]. Recently, we identified a putative SBD of the CBM20 type from *A. thaliana* GWD3 [25] (Fig. 3), and starch binding capacity has been demonstrated by *in vitro* binding assay with *A. thaliana* protein extract [26].

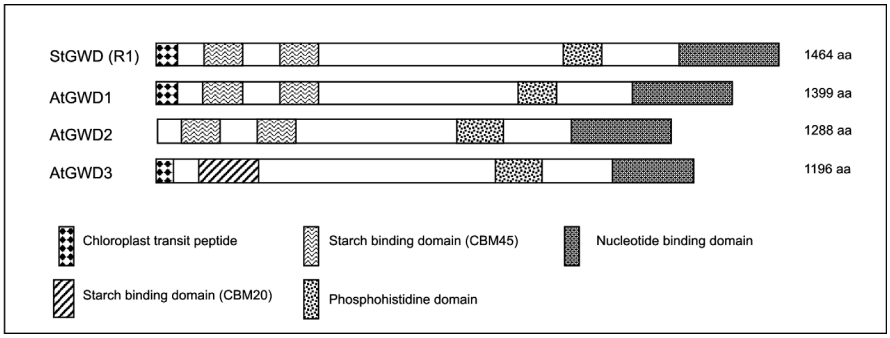


Figure 3. SBDs in plant starch dikinases. The putative domain structure (SBDs in zig-zag and diagonals) of the three GWD homologues found in the *Arabidopsis thaliana* genome and the one found in *Solanum tuberosum*. The sizes of the modules are only approximate. The sizes of the full length enzymes in amino acid residues (aa) are indicated.

In the present study, the isolated CBM20 from GWD3 was cloned and produced in *Escherichia coli* and its binding properties were compared to the previously characterised CBM20 of GA from *A. niger* [14]. Both SBDs were expressed as separate domains and purified. Molecular models were generated for the CBM20 module in GWD3. The interactions of the SBDs with well-defined soluble substrates were probed by surface plasmon resonance (SPR) to determine binding constants, substrate preference and number of binding sites. Binding to starch granules was visualised by confocal laser scanning microscopy (CLSM). Update, there has been no report of a CBM20 SBD of plant origin. This is the first report of an isolated CBM20 module originating from an intracellular non-hydrolytic enzyme involved in starch metabolism.

MATERIALS AND METHODS

Construction of expression plasmids

Since the annotation of the GWD3-CBM was only made *in silico* based on bioinformatics, constructs with three different lengths were made (encoding amino acid sequence 68-170, 68-179 and 68-184) to address the uncertain position of the domain border. Moreover, different expression systems (*E. coli* and *P. pastoris*) and different purification tags were used to maximise the possibility of producing soluble protein. The most stable gene product (encoding amino acid sequence 68-184) was obtained as described below and the data presented in this paper is based on this construct. A pYES2.1/V5-His-TOPO plasmid (Invitrogen, Paisley, UK) containing the full-length GWD3 cDNA (GenBank accession no. AY747068) [25]

was used as a PCR template to amplify Asp^{68}-Gly^{184} with the following primers; Forward: AACATGCCATGGCAGATGGATCAGGAACGAAAGTG (*Nco*I site underlined) Reverse: AACCGCTCGAGACCAACATCATCATCATTACCAAC (*Xho*I site underlined). The PCR product was cleaved by *Nco*I and *Xho*I (New England Biolabs, Hitchin, UK) and subsequently ligated into the pET-28a(+) vector (Novagen, Madison, WI, U.S.A.). The plasmid was transformed into *E. coli* BL21a (Invitrogen, Paisley, UK) and correct inserted DNA was verified by sequencing.

Expression and purification of recombinant GWD3-SBD

The *E. coli* strain BL21a harbouring pET-28a(+) with the GWD3-SBD was grown at 37ºC in LB medium with 50 µg/ml kanamycin to an optical density of 0.5 at 600 nm. Isopropyl β-D-thiogalactopyranoside was added to a final concentration of 250 µM and the culture was incubated at 20ºC for 20 hours. The cells were pelleted by centrifugation at 12,000 *g* for 20 min at 4ºC, re-suspended in 20 ml buffer A (20 mM sodium phosphate, 0.5 M NaCl, 10 mM imidazole, pH 8) and disrupted by sonication. The extracts were clarified by centrifugation at 12000 x g for 30 min at 4ºC. The sample was loaded on a 1 ml Ni-NTA-column (QIAgen, Ballerup, Denmark), washed with 15 column volumes and eluted with 60 % buffer B (20 mM sodium phosphate, 0.5 M NaCl, 500 mM imidazole, and 10% glycerol, pH 8). After this first purification step the GWD3-SBD was ~50-70 % pure (Fig. 4), therefore the eluted fraction was applied to a 6 ml β-cyclodextrin (β-CD) linked Sepharose column for a second purification step. Bound protein was washed with 2 column volumes buffer (50 mM NaCl, 250 mM NaCl pH 7.5) and eluted in 50 mM Hepes, 10 mM β-CD, pH 7.5. Fractions were analysed by Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) [27] and combined for buffer exchange by dialysis against 50 mM Hepes, pH 7.5 (molecular weight cut-off, MWCO 3.5) and stored at 4 ºC. Protein concentration in the enzyme preparation was measured spectrophotometrically at 280 nm using the theoretically calculated molar extinction coefficient (ε=28810 M^{-1} cm^{-1}). SDS-PAGE analysis was carried out to assess purity (Fig. 4). The GA-SBD from *A. niger* was kindly provided by Novozymes (Bagsvaerd, Denmark). Expression was carried out as described by [28] and purified using starch affinity chromatography.

Homology modelling

A three dimensional model of GWD3-SBD was built by homology modelling using Modeler v6.1 [29]. The SBD of GA from A. niger (PDB

code 1KUL, 26% identity and 57.7% similarity) residues 509-616 was used as template. Residues 170-184 from the GWD3-SBD were deleted from the model structure, since there were no corresponding residues in the *A. niger* GA-SBD. Because of the low homology to the template, the structure was relaxed by molecular dynamics simulation, using the program NAMD [30].

Differential scanning calorimetry (DSC)

The stability of the SBD was assessed by differential scanning calorimetry using a VP-DSC calorimeter (MicroCal, Northampton, MA) with a cell volume of 0.52 ml [31]. Protein solutions with a concentration of 2 mg/mL in 50 mM Hepes, pH 8.0 were used. Protein samples were spun down at 4 °C for 40 min at 14000 g and sterile filtered prior to the run. Buffer from the last dialysis step was sterile filtered and used to collect blank scans. Samples and buffers were degassed by vacuum at 15 °C for 15 min. All scans were collected between 15 to 100 °C at a rate of 1 °C/min. Samples were pre-temperated for 10 min at 15 °C. In order to remove residual β-CD present after protein purification, purified protein was subjected to an additional His-trap purification step and rescanned as above following buffer exchange.

Surface Plasmon Resonance (SPR) analysis

SPR was performed using a Biacore T100 (GE healthcare, Uppsala, Sweden). The SBDs were biotinylated in the presence of 10 mM β-CD, to block the carbohydrate binding sites, by the reagent EZ-Link1 Sulfo-NHS-LC-Biotin (Pierce, Rockford, Il, U.S.A.). The reagent was added to 100 µl GWD3-SBD (1 mg/ml, in a 50 mM Hepes pH 7.5 buffer) in a final 10-fold molar excess in a total volume of 120 µl. The reaction mixture was incubated for 1 hour at 15 ºC and the labelled GWD3-SBD was desalted twice on a 0.5 ml Zeba Desalt Spin Column (Pierce, Rockford, Il, U.S.A). Biotin-GWD3-SBD was immobilized on a streptavidin-coated biosensor chip (Sensor Chip SA) according to the manufacturer's instructions. Carbohydrate ligands were dissolved in running buffer (50 mM Hepes pH 7.0 or 50 mM sodium acetate pH 5.5) including 0.5% P20 surfactant (GE healthcare, Uppsala, Sweden). Binding of α-, β- and γ-CD was analysed at 15 different concentrations in the range 0.01−10 mM. α-and γ-CD were analysed in duplicates and β-CD using four replicates. Binding experiments were performed at 10 °C both in the sample compartment and in the SPR cells at a flow rate of 30 µl/min. The apparent equilibrium dissociation constant (K_d), was determined by the

BIAevalution version 1.1 software (GE healthcare, Uppsala, Sweden) and Sigma Plot 9.0.

Confocal Laser Scanning Microscopy (CLSM)

SBD was labelled with fluorescein 5-EX succinimidyl ester (Invitrogen, Paisley, UK) by amine modification. Fluorescein 5-EX succinimidyl ester was dissolved in dimethyl formamide at 10 mg/ml immediately before starting the reaction. The reagent was added to 100 µl GWD3-SBD (1 mg/ml, in a 50 mM Hepes, 10 mM β-CD buffer adjusted to pH 8 by 0.2 M sodium carbonate) in a final 4-fold molar excess in a total volume of 140 µl. The reaction mixture was incubated for 1 hour at 20 ºC with continuous stirring.

The fluorescein conjugated proteins were separated from excess reagent by desalting twice on a 0.5 ml Zeba Desalt Spin Column (Pierce, Rockford, Il, U.S.A). Labelled protein binding to starch granules was visualised by confocal laser scanning microscopy (TCS SP2, Leica Microsystems, Germany) as previously described [32]. For fluorescein, a 488 nm laser was used for excitation, and emission of light was detected between 500 and 550 nm. A laser power of 25% was maintained during acquisition of all images, and the gain was varied to prevent saturation of the detector and to ensure comparable intensities. The objectives used were HC PL APO CS 20.0x0.70 IMM/COR and HCX PL APO 63.0x1.20 W CORR UV. Image analysis was performed with the TCS SP2 software.

RESULTS AND DISCUSSION

The structure of GWD3-SBD assessed by homology modelling

The amino acid sequence of the putative GWD3-SBD [25] was used in a BLAST search to identify a template for homology modelling. The deduced amino acid sequence of the CBM20 from *A. thaliana* shows relatively low identity to other CBM20s. *A. niger* GA SBD (PDB code: 1KUL) displayed the highest sequence identity (26% identity and 57.7% similarity) in the Protein Data Bank. The relatively low sequence identity between target and template proteins rendered the GWD3-SBD modelling a challenging task. However, it has been revealed that CBM20s both in eukaryotes and prokaryotes share a similar structural fold [5].

The modelled three-dimensional structure of GWD3-SBD exhibited a β-sandwich fold with an immunoglobulin-like topology (Fig. 2, right). In

total, seven β-strands are distributed into two β-sheets. The architecture of the ligand-binding sites are structurally well conserved among the CBM20s. Both tryptophans in binding site 1 (W543 and W590 in GA-SBD; W103 and W143 in GWD3-SBD) are conserved at sequence level, and possibly at the structural level as well, as both resides are superimposable between the GA structure and the homology model of the GWD. Binding site 2 is also represented by the two aromatic amino acids, F87 and W111 in the GWD3-SBD as compared with Y527 and aW563 in the GA SBD structure.

Production of recombinant GWD3-SBD

The length of GWD3-SBD was defined by using the *A. niger* glucoamylase SBD as template and resulted in a 102 amino acid long polypeptide (aa residues 68-170). This sequence tagged by a C or N-terminal hexa-His peptide or by GST, resulted in low production yields in *E. coli* and *P. pastoris*. Furthermore, binding activity of purified proteins diminished within days. Further analysis of the primary structure suggested that the GWD3-SBD might have a longer C-terminus as judged by the presence of hydrophobic residues downstream of the annotated sequence and the lack of an apparent secondary structure shift such as the presence of glycines or prolines. Two additional constructs encoding amino acid sequence 68-179 and 68-184 were prepared. The most stable product corresponded to the latter sequence and was prepared in *E. coli* fused to a C-terminal His tag. Based on conserved amino acid sequences shared by other CBM20s, this sequence is longer than expected, but gave a relatively more stable protein. Analysis by SDS–PAGE of soluble and insoluble protein fractions, obtained after sonication and centrifugation, showed that most of the recombinant protein was in the soluble fraction. GWD3-SBD was purified to apparent homogeneity using a two step procedure including initial capture by His-tag affinity chromatography which resulted in 50–70 % pure fractions based on SDS-PAGE. As a second purification step, β-CD-Sepharose affinity chromatography was used. After elution with 10 mM β-CD GWD-SBD was electrophoretically pure (Fig. 4). The molecular mass of the protein as calculated from the amino acid sequence is 14 kDa, including the His_6 tag at the C-terminus. This is in agreement with its apparent molecular mass as estimated by SDS-PAGE analysis. Under optimised conditions the overall yield of pure GWD3-SBD from a 3-litre *E. coli* cultivation was typically 2 mg.

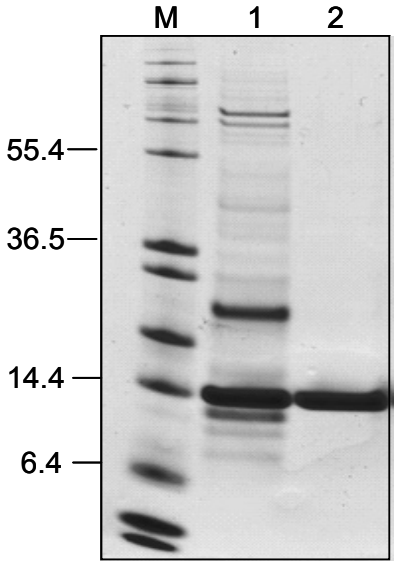


Figure 4. SDS-PAGE analysis of purified GWD3-SBD. Lane M: molecular weight standard (kDa), Lane 1: GWD3-SBD after His-trap purification, Lane 2: GWD3-SBD after β-CD purification.

Thermal stability

The conformational stability of the recombinant GWD3-SBD was evaluated by DSC. The calorimetric traces of samples obtained after the β-CD affinity purification and extensive dialysis showed a broad endothermic transition with an assigned T_m of 44.6 ºC (Fig. 5). After this first transition, the trace was irregular showing a broad transition between 50–75 ºC. The qualitative appearance and the T_m value of the first transition were reproducible within 0.6 ºC, but the second transition was not reproducible between different samples. Samples which were re-purified on a His-trap column to remove any traces of β-CD displayed significantly smaller enthalpy change, and had a single broad transition with a T_m roughly at 40 ºC. All thermal transitions were completely irreversible as judged by the lack of recovery of peak area after cooling and rescanning. These results suggest that the first thermal transition represents the thermal unfolding of the GWD SBD followed by aggregation of the unfolded protein. A second observed peak was only found in the presence of trace amounts of β-CD remaining from the purification procedure. These transitions were not visible in the re-purified sample. These results also suggest that β-CD binding increases the thermal stability of the domain as the re-purified samples showed roughly 5 ºC lower T_m and significantly smaller enthalpy change.

The small enthalpy change during the unfolding transition and the broadness of the peak are suggestive of the low conformational stability of the GWD3-SBD when present in isolated form. For example the GA-SBD from *A. niger* has a melting temperature of 56.7 ± 0.2ºC and is characterised by reversible unfolding – refolding [33]. It is likely that the reason for the instability of the GWD-SBD is caused by the fact that the domain is expressed as an isolated domain and gets destabilised by exposure of a hydrophobic surface and lack of interactions between the GWD3-SBD and the rest of the protein.

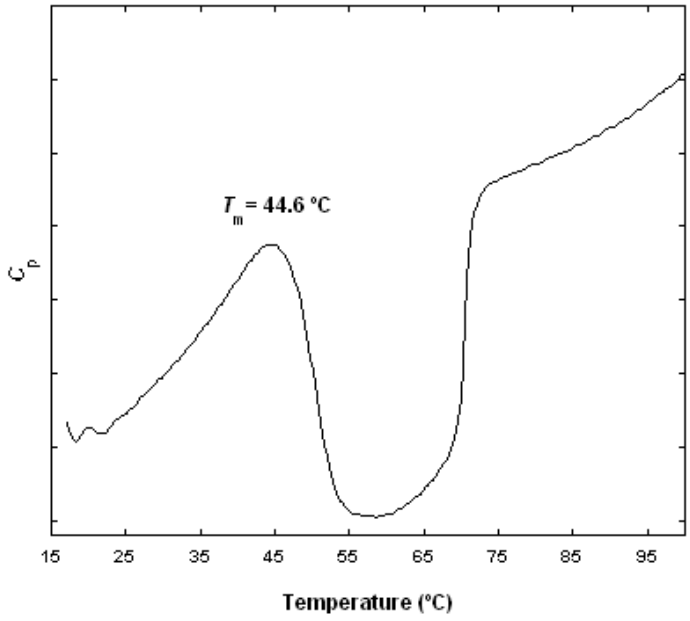


Figure 5. The calorimetric scan of the GWD3-SBD. The baseline subtracted calorimetric GWD3-SBD scan was collected at a scan rate of 1 ºC/min and with a protein conc. is 2 mg/ml. The units are omitted from the Y-axis due to the irreversibility of the scan precluding a quantitative thermodynamic treatment.

Affinity to cyclodextrins (CD) as determined by Surface Plasmon Resonance (SPR)

The affinity of α-, β-, and γ-CD, composed of 6, 7 and 8 α-1,4 linked glucose residues, respectively, was investigated using SPR (Table 1). The two domains, GWD3-SBD and GA-SBD were immobilized on sensor chips to levels of approximately 300 response units (RU) using the standard amine coupling method. CDs were injected over the immobilized surface at concentrations ranging from 10 μM to 10 mM. To keep the GWD3-SBD protein stable for longer periods the temperature was set to 10 ºC during the SPR experiments. As compared with data collected at 25 ºC, the K_d values were very similar (data not

shown) but the binding levels were maintained at high levels for several hours. The physiological pH in the chloroplast ranges from 6.5–8 and therefore the affinity was analysed at pH 7.0. Interestingly, K_d values of this domain for β-CD are in the mM range compared to the μM range reported in other CBM20 [14,34]. To compare the GA-SBD and the GWD3-SBD affinity towards β-CD affinity was determined also at pH 5.5 (the pH for optimal binding of GA-SBD) and at this pH the GWD3-SBD had invariant affinity as compared to pH and 70-fold higher K_d than the GA SBD. The different GWD3-SBD constructs expressed in either E. coli or P. pastoris were tested for β-CD affinity at pH 7.0 and all gave K_d values in the same range as the present construct (Table 1). This shows that even expressed in different hosts the GWD3-SBDs affinity towards β-CD was similar and possible glycosylation of the protein generated in P. pastoris did not significantly affect the affinity of the domain.

To compare with previously determined affinity values GA-SBD data were fitted to a one site model and provided an affinity to β-CD of K_d = 7.5 μM. This value is consistent with data previously obtained by UV difference spectroscopy, which gave a K_d = 14.4 μM [14] and by NMR titration which gave a K_d = 9 μM [34], respectively. The functional difference between the two sites of the GA-SBD was assessed by constructing two mutants, W563K and W590K [14]. The K_d values obtained for the mutants were found to be 28 and 6.4 μM for W563K and W590K, respectively. The results obtained by SPR in this study agree with the previously reported K_d values, when the data is fitted to a two binding site model. The two dissociation constants were calculated to be K_{d1} = 2.0 μM and K_{d2} = 31.3 μM. This clearly shows that the GWD-SBD binds the starch mimic β-CD with an affinity roughly two orders of magnitude weaker than the typical CBM20 SBD occurring in the context of extracellular starch degradation [14]. The SPR data obtained on the GA domain validate the experimental set up, and along with the fact that the different GWD constructs had similarly weak affinities preclude experimental artefacts being responsible to these differences.

Table 1: Dissociation constants for GWD3-SBD and GA-SBD for different CDs determined using SPR and calculated using a one site model.

CBM20	**Ligand**	**pH**	$\boldsymbol{K_d}$ (mM)
GWD3-SBD	α-CD	7.0	0.5±0.1
GWD3-SBD	β-CD	7.0	1.2±0.1
GWD3-SBD	γ-CD	7.0	2.0±0.1
GWD3-SBD	β-CD	5.5	0.5±0.1
GA-SBD	β-CD	5.5	0.0075±0.0001

Binding of GWD3-SBD to starch granules

Binding of fluorescent labelled GWD3-SBD to maize endosperm starch granules was visualised qualitatively by CLSM. A molar labelling degree of 0.2-0.5 was obtained with Fluorescein 5-EX succinimidyl ester. The maize starch granules were incubated with the labelled GWD3-SBD for 45 minutes at 4 ºC with gentle rotation. Clear fluorescence was detected at the surface of the starch granules (Fig. 6). Moreover, SBD binding in channels and cavities of the starch granules was observed. No detectable background fluorescence was observed. As a control, either BSA labelled with Fluorescein 5-EX succinimidyl ester or the dye alone was added to starch granules. No fluorescence on starch granules was observed in these samples.

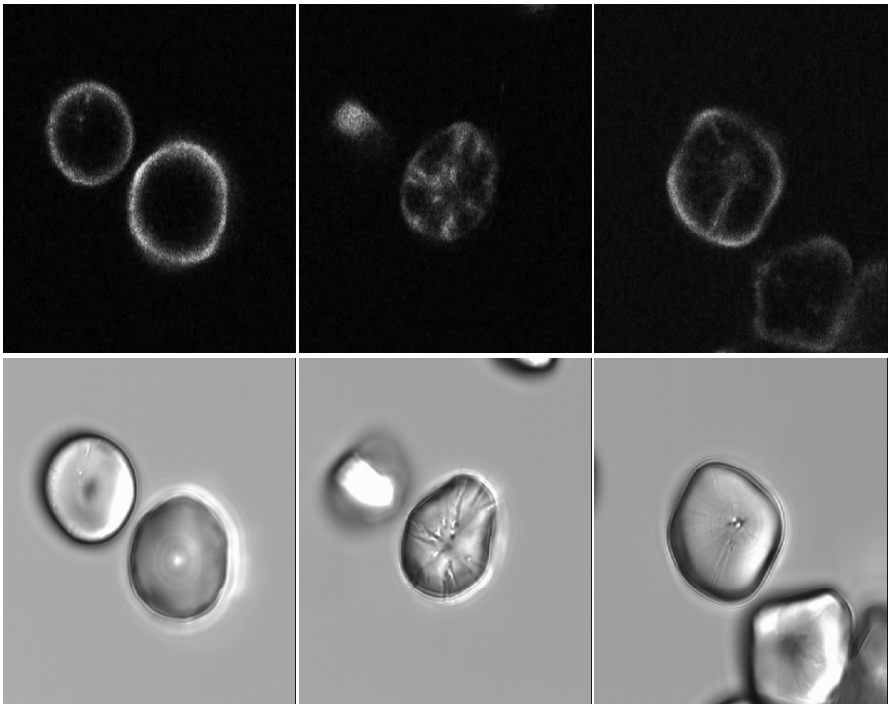


Figure 6. CLSM image of maize starch granules incubated with fluorescein-labelled CBM20 from GWD3 (top) demonstrating its affinity for starch granules. Brightfield image of the same granules (bottom).

CONCLUDING REMARKS

The present data provides the first functional characterisation of a plant CBM20. The homology model of GWD3-SBD serves as a guideline for design of site-directed mutants. The relative locations of the aromatic residues in the substrate binding sites are conserved and they form a surface exposed hydrophobic platform appropriate for binding starch. The reason for the low thermal stability of the GWD3-SBD as compared to the *A. niger* GA-SBD is not known. One explanation could be that the domain interacts tightly with the adjacent enzyme domain(s) causing loss of stability when separated from the catalytic domain. There is no linker separating the SBD from the catalytic domain, as is found in GA. A 70-fold lower K_d was found for the GWD3-SBD compared to the GA-SBD using β-CD as ligand, identifying the GWD3-SBD as a low affinity SBD together with the CBM45 found in GWD1. It was, furthermore, demonstrated that the module binds to native starch granules.

The low affinity of the present plant CBMs is anticipated to have a physiological function. The CBM45 in GWD1 has also been reported to be a low affinity SBD. It is predominantly found in soluble form when extracted from illuminated plants, but it was found bound to the starch granules in the dark [1,35]. Such detailed information does not exist for GWD3, but the low affinity value for the GWD3-SBD at physiologically relevant pH, suggests that binding might be reversible and might be redox regulated like for GWD1 [1]. Interestingly, intracellular CBMs in this study present in dikinases display substantially lower affinity than the CBMs attached to hydrolases that are mostly extracellular. Further biochemical and structural studies are required to investigate the biological significance of this lower affinity binding to starch and to identify structural elements responsible for the large affinity difference within family 20 CBMs.

ACKNOWLEDGEMENTS

This work was supported by the Danish Research Council for Natural Science, the Danish Research Council for Technology and Production Sciences, the Carlsberg Foundation, a H.C. Ørsted Post-doctoral Fellowship from DTU (MJB) and a FOBI PhD stipend (CC).

REFERENCES

1. Mikkelsen,R., Mutenda,K.E., Mant,A., Schurmann,P., Blennow,A. Proc. Natl. Acad. Sci. U.S.A. 102, 1785 (2005).
2. Blennow,A., Viksø-Nielsen,A., Morell,M.K. Eur. J. Biochem. 252, 331 (1998).
3. Denyer,K., Johnson,P., Zeeman,S., Smith,A.M. J. Plant. Phys. 158, 479 (2001).
4. Machovic,M., Janecek,S. Cell Mol. Life Sci. 63, 2710 (2006).
5. Rodriguez-Sanoja,R., Oviedo,N., Sanchez,S. Curr. Opin. Microbiol. 8, 260 (2005).
6. Lawson,C.L., van,M.R., Strokopytov,B., Rozeboom,H.J., Kalk,K.H., de Vries,G.E., Penninga,D., Dijkhuizen,L., Dijkstra,B.W.. J. Mol. Biol. 236, 590 (1994).
7. Svensson,B., Larsen,K., Svendsen,I., Boel,E. Carlsberg Res. Commun. 48, 529 (1983).
8. Boraston,A.B., Bolam,D.N., Gilbert,H.J., Davies,G.J. Biochem. J. 382, 769 (2004).
9. Southall,S.M., Simpson,P.J., Gilbert,H.J., Williamson,G., Williamson,M.P. FEBS Lett. 447, 58 (1999).
10. Juge,N., Nohr,J., Le Gal-Coeffet,M.F., Kramhoft,B., Furniss,C.S., Planchot,V., Archer,D.B., Williamson,G., Svensson,B. Biochim. Biophys. Acta 1764, 275 (2006).
11. Morris,V.J., Gunning,A.P., Faulds,C.B., Williamson,G., Svensson,B. Starch-Starke 57, 1 (2005).
12. Sorimachi,K., Le Gal-Coeffet,M.F., Williamson,G., Archer,D.B., Williamson,M.P. Structure 5, 647(1997).
13. Giardina,T., Gunning,A.P., Juge,N., Faulds,C.B., Furniss,C.S.M., Svensson,B., Morris,V.J., Williamson,G. J. Mol. Biol. 313, 1149 (2001).
14. Williamson,M.P., Le Gal-Coeffet,M.F., Sorimachi,K., Furniss,C.S., Archer,D.B., Williamson,G. Biochemistry 36, 7535 (1997).
15. Li,J.H., Vasanthan,T., Hoover,R., Rossnagel,B.G. Food Chem. 84, 621 (2004).
16. Viksø-Nielsen,A., Andersen,C., Hoff,T., Pedersen S. Biocat. Biotransform. 24, 121 (2006).
17. Svensson,B., Jespersen,H., Sierks,M.R., MacGregor,E.A. Biochem. J.264, 309 (1989).
18. Janecek,S., Sevcik,J. Febs Lett. 456, 119 (1999).
19. Sorimachi,K., Jacks,A.J., Le Gal-Coeffet,M.F., Williamson,G., Archer,D.B., Williamson,M.P. J. Mol. Biol. 259, 970 (1996).
20. Jacks,A.J., Sorimachi,K., Le Gal-Coeffet,M.F., Williamson,G., Archer,D.B., Williamson,M.P. Eur. J. Biochem. 233, 568 (1995).
21. Penninga,D., van der Veen,B.A., Knegtel,R.M.A., van Hijum,S.A.F.T.,

Rozeboom,H.J., Kalk,K.H., Dijkstra,B.W., Dijkhuizen,L. J. Biol. Chem. 271, 32777 (1996).

22. Mikkelsen,R., Suszkiewicz,K., Blennow,A. Biochemistry 45, 4674 (2006).
23. Machovic,M., Svensson,B., MacGregor,E.A., Janecek,S. FEBS J. 272, 5497 (2005).
24. Paldi,T., Levy,I., Shoseyov,O. 372, 905 (2003).
25. Baunsgaard,L., Lütken,H., Mikkelsen,R., Glaring,M.A., Pham,T.T., Blennow,A. Plant J. 41, 595 (2005).
26. Kötting,O., Pusch,K., Tiessen,A., Geigenberger,P., Steup,M., Ritte,G. Plant Physiol 137, 242 (2005).
27. Laemmli,U.K. Nature 227, 680 (1970).
28. Christensen,T., Woeldike,H., Boel,E., Mortensen,S.B., Hjortshøj,K., Thim,L., Hansen,M.T. Bio Technol 6, 1419 (1988).
29. Sali,A., Blundell,T.L. J. Mol. Biol. 234, 779 (1993).
30. Kalé,K., Skeel,R., Bhandarkar,M., Brunner,R., Gursoy,A., Krawetz,N., Phillips,J., Shinozaki,A., Varadarajan,K., Schulten,K. J. Comput. Phys. 151, 283 (1999).
31. Plotnikov,V.V., Brandts,J.M., Lin,L.N., Brandts,J.F. Anal. Biochem. 250, 237 (1997).
32. Blennow,A., Hansen,M., Schulz,A., Jorgensen,K., Donald,A.M., Sanderson,J. J. Struct. Biol. 143, 229 (2003).
33. Christensen,T., Svensson,B., Sigurskjold,B.W. Biochemistry 38, 6300 (1999).
34. Le Gal-Coeffet,M.F., Jacks,A.J., Sorimachi,K., Williamson,M.P., Williamson,G., Archer,D.B. Eur. J. Biochem. 233, 561 (1995).
35. Ritte,G., Lorberth,R., Steup,M. Plant J. 21, 387 (2000).

PAPER II

VERSITA

Biologia **63**/6: 967—979, 2008
Section Cellular and Molecular Biology
DOI: 10.2478/s11756-008-0164-2

REVIEW

An enzyme family reunion – similarities, differences and eccentricities in actions on α-glucans

Eun-Seong SEO[1], Camilla CHRISTIANSEN[1,2], Maher ABOU HACHEM[1], Morten M. NIELSEN[1], Kenji FUKUDA[1,3], Sophie BOZONNET[1,3], Andreas BLENNOW[2], Nushin AGHAJARI[4], Richard HASER[4] & Birte SVENSSON[1]*

[1]*Enzyme and Protein Chemistry, Department of Systems Biology, Technical University of Denmark, Søltofts Plads, Bldg. 224, DK-2800 Kgs. Lyngby, Denmark; e-mail: bis@bio.dtu.dk*
[2]*Plant Biology Laboratory, Faculty of Life Sciences, University of Copenhagen, DK-1871 Frederiksberg, Denmark*
[3]*Carlsberg Laboratory, Gamle Carlsberg Vej 10, DK-2500 Valby, Denmark*
[4]*Laboratorie de BioCristallographie, Institut de Biologie et Chimie des Protéines, UMR 5086–CNRS/Université de Lyon, IFR128 "BioSciences Gerland Lyon-Sud", F-69367, Lyon, France*

Abstract: α-Glucans in general, including starch, glycogen and their derived oligosaccharides are processed by a host of more or less closely related enzymes that represent wide diversity in structure, mechanism, specificity and biological role. Sophisticated three-dimensional structures continue to emerge hand-in-hand with the gaining of novel insight in modes of action. We are witnessing the "test of time" blending with remaining questions and new relationships for these enzymes. Information from both within and outside of ALAMY_3 Symposium will provide examples on what the family contains and outline some future directions. In 2007 a quantum leap crowned the structural biology by the glucansucrase crystal structure. This initiates the disclosure of the mystery on the organisation of the multidomain structure and the "robotics mechanism" of this group of enzymes. The central issue on architecture and domain interplay in multidomain enzymes is also relevant in connection with the recent focus on carbohydrate-binding domains as well as on surface binding sites and their long underrated potential. Other questions include, how different or similar are glycoside hydrolase families 13 and 31 and is the lid finally lifted off the disguise of the starch lyase, also belonging to family 31? Is family 57 holding back secret specificities? Will the different families be sporting new "eccentric" functions, are there new families out there, and why are crystal structures of "simple" enzymes still missing? Indeed new understanding and discovery of biological roles continuously emphasize value of the collections of enzyme models, sequences, and evolutionary trees which will also be enabling advancement in design for useful and novel applications.

Key words: glycoside hydrolase families 13, 31, 57, 70, and 77; crystal structures; substrate specificities; surface binding sites; degree of multiple attack; starch granules; calcium ions; starch-binding domains; barley α-amylase.

Abbreviations: AMY1, barley α-amylase 1; AMY2, barley α-amylase 2; BASI, barley α-amylase/subtilisin inhibitor; CBM, carbohydrate-binding module; β-CD, β-cyclodextrin; DP, degree of polymerization; GBD, glucan-binding domain; GH, glycoside hydrolase; GPI, glycosylphosphatidylinositol; GWD, glucan, water dikinase; SBD, starch-binding domain.

Introduction

The group of starch-degrading and related enzymes active on α-glucosides and α-glucans belong to glycoside hydrolase families 13, 14, 15, 31, 57, 70, and 77 (http://www.cazy.org/). The very large α-amylase – or glycoside hydrolase 13 (GH13) – family represented by more than 4500 sequences in databases, is steadily growing and enzymes have emerged in bacteria, filamentous fungi, and plants which play hitherto unidentified roles in biological systems. Moreover, structural biology keeps providing new three-dimensional structures, the exceptionally impressive example being the crystallisation and solving of the structure of a GH70 member, the glucansucrase from *Lactobacillus reuteri* 180 (Pijning et al. 2008). Some enzyme newcomers are engaged in conversion of large substrates which is commonly facilitated by dedicated carbohydrate-binding domains. The whole area of protein-polysaccharide interaction and processing including the relationship between binding and catalysis and synergistic action of various enzymes develops rapidly and improves insight on the complexity of the reactions at the molecular level.

* Corresponding author

Springer

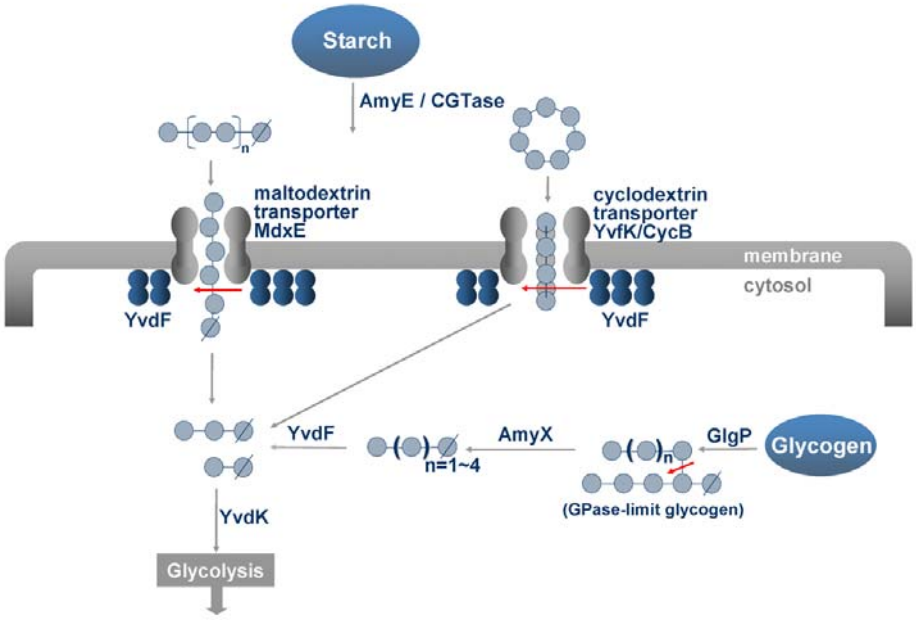


Fig. 1. Proposed model of sugar utilization in *Bacillus subtilis* (courtesy K.H. Park).

New roles and diversity of α-glucan-active enzymes in biological systems

Bacteria and fungi are well known for producing starch degrading and related enzymes that secure access to and utilization of nutrients. In recent years, however, also enzymes with key roles in intracellular processes have been identified to belong to selected sub-families of GH13 (Stam et al. 2006). This draws attention to multispecificity as one of the major problems in the genomic era for correct prediction of specificity and biological role based on a GH family assignment.

Park and co-workers (Park et al. 2008) report new roles for hydrolases and debranching enzymes engaged in sugar utilisation in *Bacillus subtilis* where genome-mining indicated glucosidases/transglucosidases and glucosyltransferases involved in degradation of maltodextrin and glycogen (Fig. 1). The *B. subtilis* mutants *yvdF* and *amyX* were defective in the carbohydrate hydrolase (YvdF) or a debranching enzyme (AmyX) and were examined *in vivo* and *in vitro*. Wild-type *B. subtilis* takes up maltoheptaose and β-cyclodextrin via two distinct transporters, MdxE and YvfK/CycB, respectively. YvdF is localized close to the cell membrane and immediately hydrolyses these sugars to give linear maltodextrins. Breakdown of glycogen by cell extracts increased in the order of wild-type > *yvdF* > *amyX* > *amyX/yvdF* mutants. The side chain length preference of debranching enzymes is important in shaping glycogen both during synthesis and degradation. The debranching enzyme specificity can be tested by incubation with branched β-cyclodextrins (Park et al. 2008). While AmyX specifically hydrolysed side chains of 3–5 glucosyl residues, the related TreX from *Sulfolobus solfataricus* showed specificity for DP 3–7 (Park et al. 2008) and GlgX (*E. coli*) for DP 3–4. Interestingly, a pullulanase from *Nostoc* exclusively hydrolyzed long side chains of DP 9–10. The results lead to the proposition of a specific debranching mechanism of glycogen breakdown in bacteria involving isoamylase-type of activity on phosphorylase limit glycogen, which is distinct from the mechanism of the glycogen debranching enzyme in yeasts and mammals.

Currently known fungal α-amylases are well-characterized extracellular enzymes classified in glycoside hydrolase subfamily GH13_1 (Stam et al. 2006). Genome-mining in *Aspergillus niger* also identified α-glucan-acting enzymes phylogenetically annotated to GH13_1, but surprisingly these contained glycosylphosphatidylinositol (GPI)-anchor sequence motifs belonging to intracellular glucanotransferases and hydrolases or they were clustered to the family GH13_5 having no previous assignments, which by cloning and recombinant enzyme production was found to contain intracellular hydrolases with low activity (van der Kaaij et al. 2007a). Homologues of these intracellular enzymes are seen in genome sequences of all filamentous fungi studied. One of the enzymes from this new group, Amy1p from *Histoplasma capsulatum* (Marion et al. 2006), has recently been functionally linked to the formation of cell wall α-glucan (Fig. 2). To study biochemical properties of the GH13_5 cluster AmyD, a homologue from *A. niger*, was overexpressed and shown to have low hydrolysing activity on starch and to produce mainly maltotriose. Moreover three genes encoded proteins with high similarity to fungal α-amylases. Remarkably these were predicted to have a GPI-anchor in distinction to α-amylases described earlier and they furthermore lacked some highly conserved amino acids of GH13. Two enzymes AgtA and AgtB prepared recombinantly showed transglycosylation activity on maltopentaose or longer donor substrates to produce new α-1,4-glucosidic bonds, thus belonging to the 4-α-glucanotransferases. The prod-

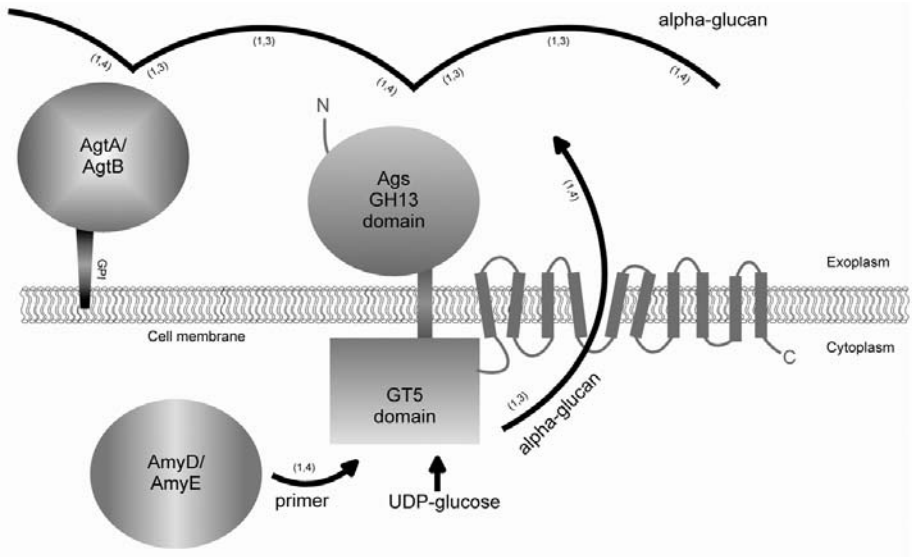


Fig. 2. Schematics of GH13 enzymes engaged in fungal cell wall biosynthesis (courtesy R.M. van der Kaaij).

ucts reached DP > 30; and small maltooligosaccharides were the most efficient acceptor substrates. AgtA, however, also used small α-1,3-linked nigerooligosaccharides as acceptor and an AgtA knockout of *A. niger* got increased susceptibility towards calcofluor white indicating defect cell walls. Homologues of AgtA and AgtB are present in other fungal species having α-glucan constituents in their cell wall (van der Kaaij et al. 2007b). Recently, also a putative α-glucosidase (AgdB) and an α-amylase (AmyC) predicted to degrade starch were reported in the *A. niger* genome (Yuan et al. 2008). Other members of GH13, GH15, and GH31 might function in alternative α-glucan modifying processes (Yuan et al. 2008).

A different type of system that holds a very high level of amylolytic activity involving an array of enzyme specificities is the germinating cereal seed (Fig. 3).

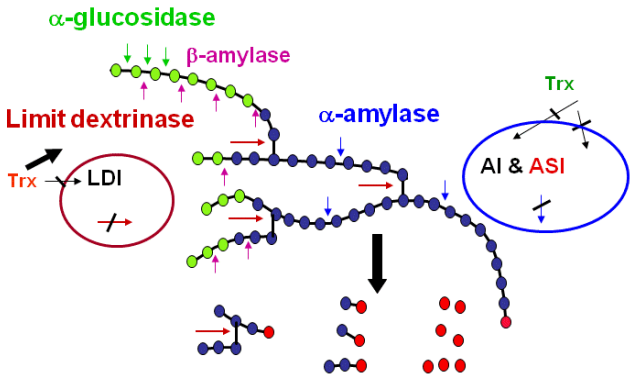


Fig. 3. Schematics of the amylolytic system in germinating cereal seeds. A segment of amylopectin is schematized and the arrows (colour code) indicate bonds hydrolysed by the different enzymes. Red spheres represent reducing ends. A bar indicates inhibition, a large black arrow indicates stimulation. LDI = limit dextrinase inhibitor; ASI = α-amylase/subtilisin inhibitor (specifically inhibiting AMY2); AI = α-amylase inhibitor (specific for exogenous enzymes); Trx = thioredoxin.

This includes numerous enzyme forms and the complexity was illustrated recently by applying a combined immunoblotting and proteomics-approach to survey molecular forms of α-amylases. The system also includes different forms of β-amylase, α-glucosidase and the debranching enzyme limit dextrinase, involved in mobilisation of endosperm starch granules. The immunoblotting of two-dimensional electrophoretic gels of aqueous extracts of germinating barley seeds developed numerous spots containing α-amylase or breakdown products thereof (Bak-Jensen et al. 2007). Among the 10 α-amylase-encoding genes in barley, four encode a member of the minor isozyme 1 (AMY1) and six a member of the major isozyme 2 (AMY2) family. Surprisingly, only one of 10 different forms identified in seven spots of varying iso-electric points (pI) containing full-length α-amylase (~45 kDa), belonged to the AMY1, while nine forms stemmed from two AMY2 subfamily members (Bak-Jensen et al. 2007). Moreover, mass spectrometry showed that all but one of 22 spots constituting different "spot trains", i.e. series of degradation products, in the range of ~20–~39 kDa were derived from one specific of the two AMY2 gene products. Only a single fragment originated from the second AMY2 encoding gene. Thus this approach not only identified the main two AMY2 genes out of six present in the genome, but it also demonstrated that one of these seems importantly more stable towards proteases in the germinating seeds. Alternatively, the gene product corresponding to the single fragment was even more sensitive to proteases and was broken down to fragments too short for detection by the two-dimensional electrophoresis. At the moment no other functional or stability difference is reported between these two members of the AMY2 subfamily. Moreover, nothing is known on the spatio-temporal occurrence in the seed of the two AMY2 gene products or for that sake where, when, or if the four other AMY2 genes are expressed in the plant. Although the α-amylase

group certainly is the most complex, several forms and degradation products were also observed for limit dextrinase, β-amylase, and α-glucosidase in extracts of germinating seeds. Remarkably, combined proteome analysis of gibberellic acid-treated aleurone layer cells and the corresponding excreted proteome in the culture liquid expected to correspond to the enzymes normally transferred from the aleurone layer into the endosperm for mobilisation of starch, showed that α-amylase was rapidly secreted, while limit dextrinase appeared very late in the culture liquid. Hence, limit dextrinase is presumably excreted late into the endosperm during germination (A. Shahpiri et al., manuscript in preparation). Recently, a chemical genetics approach has been applied by soaking germinating barley seeds with various α-glucosidase inhibitors to selectively observe the consequence of inactivating a given enzyme activity. One such inhibitor strongly affected the morphology of the roots and the acrospire of the germinating seed (Stanley et al. 2007).

The proteins associated with starch degradation in barley include two proteinaceous inhibitors, α-amylase/subtilisin inhibitor (BASI) and limit dextrinase inhibitor (LDI), specifically regulating the activity of AMY2 and limit dextrinase, respectively. The BASI-AMY2 complex has been well described using site-directed mutagenesis, crystallography, surface plasmon resonance, and activity inhibition analyses (Vallée et al. 1998; Rodenburg et al. 2000; Nielsen et al. 2003; Bønsager et al. 2005) and found to have high stability, sub-nanomolar affinity; specific residues were assigned functional roles both in enzyme and inhibitor for the complex formation. Analysis of the LDI-limit dextrinase complex is just initiated thanks to breakthroughs with successful heterologous production of both LDI and limit dextrinase (M. Vester-Christensen et al., manuscript in preparation) and data now emerge showing sub-nanomolar affinity also for this complex (J.M. Jensen et al., unpublished results).

Finally, the protein disulfide reductase thioredoxin has been proposed to regulate the amylolytic system in barley seeds by reduction of disulfide bonds in enzymes and inhibitors (Cho et al. 1999). Barley contains two thioredoxin isoforms as well as two isoforms of an NADPH-dependent thioredoxin reductase that reduce the disulfide formed in the thioredoxin active site motif CXXC after it has reduced a target protein disulfide bond (Maeda et al. 2003; Shahpiri et al. 2008). We have developed a proteomics-based procedure that allows identification of target disulfides in protein mixtures (Maeda et al. 2005; P. Hägglund et al., manuscript in preparation). Furthermore we determined the crystal structure of a trapped complex of barley thioredoxin h and BASI to identify target protein structural requirements for thioredoxin recognition (Maeda et al. 2006). A clear distinction of the roles of the two thioredoxin as well as of the two thioredoxin reductase isoforms has not been made. *In vitro*, one specific pair is up to three times as efficient as other pairs and this pair is moreover enriched in the aleurone layer during germination (Shahpiri et al. 2008). Maybe different spatio-temporal occurrence of isoforms is an important factor in efficient recycling of oxidised thioredoxin. The impact of thioredoxin on proteins targets is currently analysed in dissected tissues from germinating barley seeds using a newly developed quantitative procedure that ranks target disulfides with regard to degree of susceptibility to thioredoxin (P. Hägglund et al., manuscript in preparation).

Functional diversity of selected GH13 and GH57 members

The α-amylase family GH13, GH70 and GH77 together constitute clan GH-H (http://www.cazy.org/). GH13 is the largest of these families in terms of both the number of enzyme specificities and the number of sequence entries. Recently, the diversity within GH13 was emphasized by definition of subfamily clusters (Stam et al. 2006), some of which contained distinct specificity, e.g., for involvement in cell wall biosynthesis in fungi (see above; van der Kaaij et al. 2007a,b).

In a study of new neopullulanase-like GH13 members (neopullulanases, maltogenic amylases, cyclodextrinases) enzymatic and oligomerisation properties were described for enzymes recombinantly produced in *E. coli* and originating from six genes cloned from the thermophilic bacteria *Anoxybacillus*, *Thermoactinomyces*, and *Geobacillus* or environmental DNA of Icelandic hot springs (Turner et al. 2005; Nordberg Karlsson et al. 2008). Five of the enzymes had the typical N-terminal domain of neopullulanase-like enzymes, which is involved in dimerisation (Kim et al. 2001), while one enzyme originating from environmental DNA lacked the N-terminal domain. Three of the enzymes showed cyclodextrinase, maltogenic amylase as well as neopullulanase activity and most remarkably one of these enzymes (from *Thermoactinomyces*) possessed the characteristic N-terminal-domain, but was monomeric, even though cyclodextrin-degrading enzymes are usually dimeric or oligomeric (Park et al. 2000). Two of the enzymes lacked neopullulanase activity. Moreover, a moderately thermophilic enzyme without an N-terminal domain had no cyclodextrinase activity, but showed neopullulanase activity. Based on these results, the N-terminal domain seems to be required for cyclodextrinase activity, while oligomerization is not.

Extremophiles often harbour a different enzyme repertoire than other microorganisms with regard to both GH families and enzyme specificities. Recently, a new branching enzyme was described from *Thermococcus kodakaraensis*, which catalysed the formation of α-1,6-glucosidic linkages in glycogen and amylopectin by transfer after cleavage of an α-1,4-glucosidic bond. This is the first branching enzyme in GH57, which encompasses several other amylolytic specificities and is suggested to be a second "α-amylase family" (Zona et al. 2004; Murakami et al. 2006).

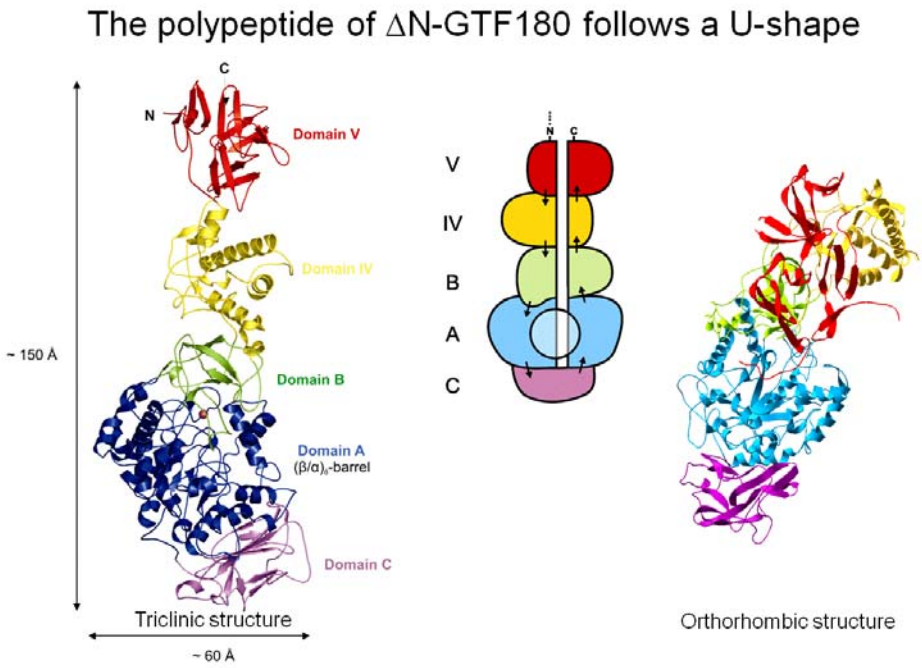


Fig. 4. Glucansucrase GTF180 three-dimensional structure and domain architecture (courtesy B.W. Dijkstra).

Break-throughs on structures of α-glucan-active enzymes

GH70 enzymes present an enormous challenge to crystallographers due their huge size and multidomain architecture. Furthermore, the earlier prediction that GH70 members have a permuted GH13 catalytic $(\beta/\alpha)_8$-barrel domain (MacGregor et al. 1996) makes it particularly exciting to get access to a three-dimensional structure. Indeed the solved structure of *Lactobacillus reuterii* 180 glucansucrase reveals an intriguing architecture in which several domains are composed of interacting segments from distant parts of the long polypeptide chain according to a U-shaped topology (Fig. 4). Numerous questions will be enlightened thanks to this structure, e.g. with regard to structure/specificity relationships in GH70 in conjunction with multiple sequence alignments revealing distinct characteristics at the conserved GH-H active-site sequence motifs (MacGregor et al. 2001). In fact, such motifs were already exploited for semi-rational manipulation of bond-type specificity in GH70 (see below). Another question is the dynamics of the GH70 molecule and how conformational changes and domain positioning accompany individual steps of the catalytic process. Thus the initial structure analysis resulted in two conformational states (Fig. 4) with domain V being mobile to swing and adapt two different positions in the global structure (T. Pijning et al., manuscript in preparation).

After extensive efforts, crystal structures were solved of several α-glycosidases from GH31. The first structure to be determined was of an enzyme encoded by an ORF from *Escherichia coli* that turned out to be an α-xylosidase, which is a less common specificity in GH31 (Kitamura et al. 2005; Lovering et al. 2005). Guided by the structure this enzyme (YicI) was engineered into an α-glucosidase (Okuyama et al. 2006) demonstrating the close relationship between the GH31 specificities. Shortly after, the structure of α-glucosidase MalA from *Sulfolobus solfataricus* was solved (Ernst et al. 2006). Finally, thanks to the recent structure of a starch lyase of GH31 (B.W. Dikstra & S. Yu, personal communication) insight into specificity determinants of this enzyme family will expand. The structural information complements kinetics analysis contributing to explain the structural basis for the different reaction mechanisms of the starch lyases and the hydrolases (Lee et al. 2003). A bootstrap diagram (Fig. 5) assigns these three different enzymes to each of three clusters, a fourth cluster contains archaeal α-xylosidases (Ernst et al. 2006). Among the "oldest" enzymes in this family are the sucrase-isomaltase and maltase-glucoamylase both from the intestinal brush border and each composed of two GH31 members originating from gene duplication. Very recently the structure of the N-terminal subunit of the human maltase-glucoamylase was determined (Sim et al. 2008) and found to represent the poorly inhibited maltase activity, whereas the C-terminal subunit has the higher catalytic activity (Quezada-Cavillo et al. 2008). Although its catalytic machinery is different, family GH31 has still some sequence similarity to clan GH-H at β3, β4, β7, and β8 of the catalytic $(\beta/\alpha)_8$-barrel. It has closest resemblance with the GH77 members (Janecek et al. 2007).

A relatively new GH13 member is dextran glucosidase from *Streptococcus mutans* that hydrolyses α-1,6-linkages at the non-reducing ends of dextrin and isomaltooligosaccharides (Saburi et al. 2006). The structures of the free and oligosaccharide binding form were

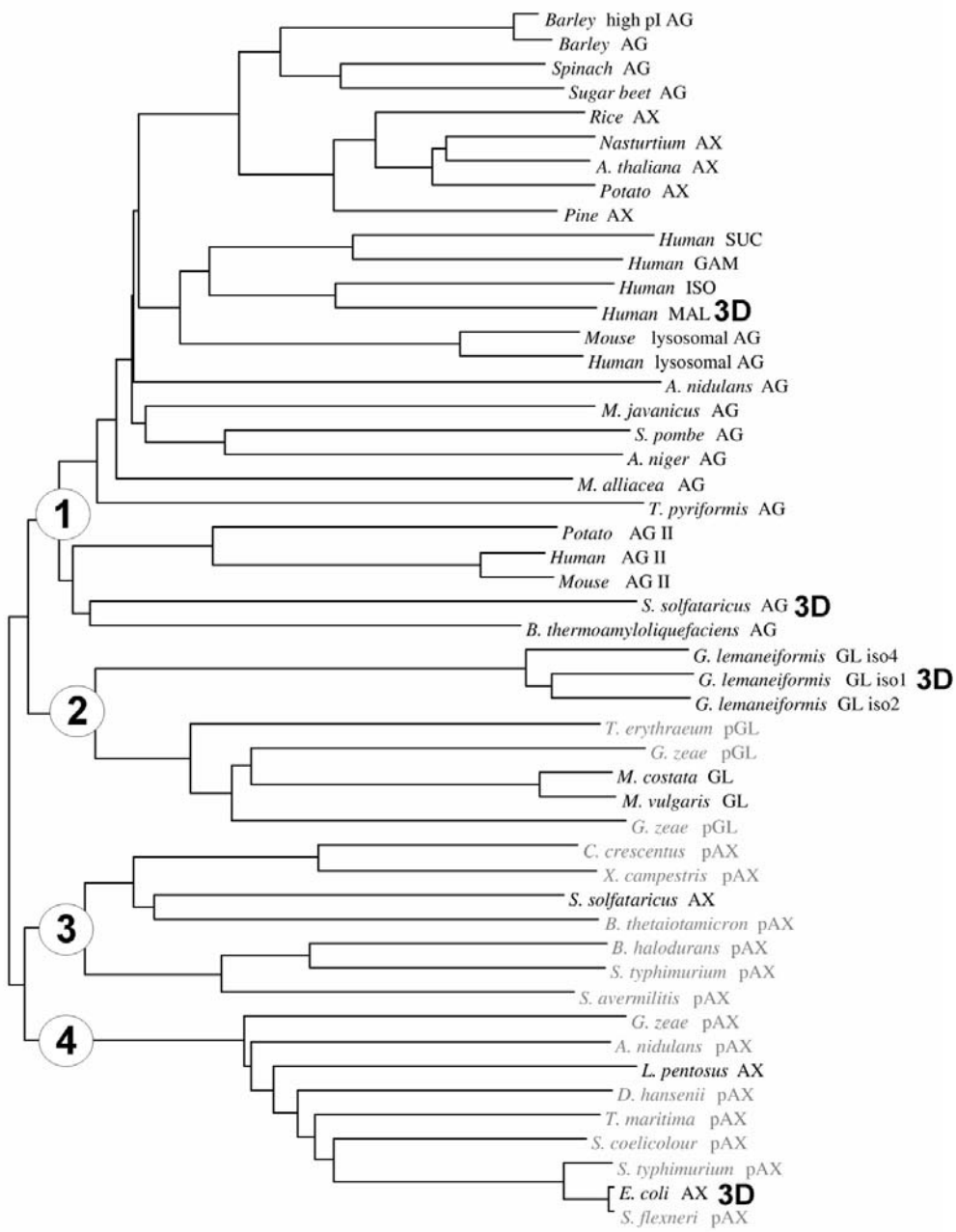


Fig. 5. Bootstrap diagram of the GH31 family. Subgroups 1–4 are indicated as are the four enzymes for which a crystal structure was determined (courtesy H.A. Ernst and L. Lo Leggio).

solved (Hondoh et al. 2008) representing a very significant advancement of knowledge as the closest relative, oligo-1,6-glucosidase of GH13, was structure determined only in its free form. Substitution of the catalytic aspartate nucleophile of the dextran glucosidase to a cysteine and subsequent oxidation to sulphinic acid improved the transglycosylation capacity of this enzyme (Saburi et al. 2007).

New structures were also presented for different starch-binding domains (SBDs). SBD of carbohydrate-binding module (CBM) family 41 uses stacking interactions for carbohydrate binding (van Bueren & Boraston 2007) and generally binding energy imparted by substantial van der Waal's interaction between complementary surfaces of sugar and CBM is supplemented by a few hydrogen bonds. It is suggested that this mode of interaction can be an evolutionary theme among non-catalytic-binding domains (Abbott et al. 2007). For the well known CBM21 family of SBDs the three-dimensional structure showed binding sites of similar interaction mode as CBM20 (Liu et al. 2007). This is in agreement with the earlier indicated relationship by a sequence alignment of the *Rhizopus oryzae* SBD with SBDs of *A. niger* glucoamylase and bacterial cyclodextrin glucanotransferases (Svensson et al. 1989). An interesting dimer of CBM21 of glucoamylase from *Rhizopus oryzae* was held together by a β-cyclodextrin molecule bridging two SBDs by binding to one of the two binding sites present in each of the SBDs (Liu et al. 2007). The bioinformatics of the relation between

CBM20 and CBM21 was recently analyzed. The evolutionary tree based on a common alignment of sequences of both modules showed that the CBM21 SBDs from α-amylases and glucoamylases are the closest relatives to the CBM20 counterparts, with the CBM20 modules from the GH13 amylopullulanases being possible candidates for the intermediate between the two CBM families (Machovic et al. 2005).

Specificity engineering in clan GH-H

The huge amount of sequence information available for clan GH-H combined with three-dimensional structures covering a broad variety of enzyme specificities (http://www.cazy.org; MacGregor et al. 2001) motivated engineering of enzymatic properties *via* various semi-rational approaches, e.g. one-dimensional/three-dimensional comparison.

Classically, specificity engineering modified the product composition for cyclodextrin glucanotransferases, neopullulanases, and maltogenic α-amylase, by taking advantage of insight into the conserved sequence motifs extending at four active site β-strands (Kuriki et al. 1996; Beier et al. 2000; MacGregor et al. 2001; Leemhuis et al. 2003). Recently, engineering of GH70 members as guided by sequences of enzymes with assigned product bond-type specificity by multiple mutational substitutions in conserved sequence motifs of GH70 succeeded to alter the α-glucan product bonds of reuteransucrase to be mainly of α-1,6- rather than α-1,4-glucosidic linkage specificity (Kralj et al. 2005). The same strategy led to enrichment of α-1,4-glucosidic linkages in the product from glucansucrase GTF180 of *Lactobacillus reuteri*. The potential for tailoring α-glucan polymer structures is enormous as is the potential towards functional design of such polymers to achieve properties adapted to specific applications (Kralj et al. 2006; van Leeuven et al. 2008).

The DSR-E glucansucrase from *L. mesenteroides* NRRL B-1299 is a unique GH70 member able to synthesize polymers containing both α-1,6- and α-1,2-glucosidic linkages. It is the largest glucansucrase (313 kDa) and has two catalytic domains, CD1 and CD2 of GH70 connected by a glucan-binding domain (GBD) (Bozonnet et al. 2002). Dissection of DSR-E revealed CD1 and CD2 to be responsible for synthesis of α-1,6- and α-1,2-linkages, respectively (Fabre et al. 2005). The truncated variant GBD-CD2 was found from the donor sucrose to be purely catalysing α-1,2-transglucosylation to dextran and α-1,6-glucooligosaccharide acceptors. Kinetic analysis revealed that the transglucosylation reaction follows a ping-pong bi-bi model ($k_{cat} = 460\ s^{-1}$) and competes with a weak sucrose hydrolase activity ($k_{cat} = 46\ s^{-1}$). By adjusting the reaction conditions, a nice panel of α-1,2-branched dextrans harbouring different and controlled degrees of branching can be synthesized (Brisson et al. 2007).

For the more subtle part, examples of site-directed mutagenesis of a single residue at one of the substrate binding subsites of barley α-amylase could change – without lowering the wild-type activity level – the relative preference for starch over oligosaccharide by a factor of 150- or contrarily caused a 50-fold preference for oligosaccharide over starch (Gottschalk et al. 2001; Mori et al. 2001; Bak-Jensen et al. 2004). Some of these mutants located at the outer subsites -6 (Y105A) or $+4$ (T212W) also elicited dramatic changes of the subsite affinity profile, thus substitution at subsite -6 of a tyrosine critical for oligosaccharide hydrolysis was accompanied by highly suppressed activity on oligosaccharides and in fact by enhanced activity on insoluble starch. Compared to wild-type AMY1 Y105A showed reduced substrate binding energy at subsite -6 of 40% and enhanced affinity for subsites -2 and $+2$ of 115% and 200%, respectively (Kandra et al. 2006). Mutation at both extreme subsites -6 and $+4$ in fact gave higher affinity at subsite $+2$ than in any of the constituent single position mutants (Kandra et al. 2006). Such insight into the impact of the subsite structure on the affinity profile provides an important tool in rational product profiling.

Impact of secondary binding sites on function

Secondary binding sites are situated outside of the substrate binding cleft in several carbohydrate-active enzymes and there is a need for understanding how such sites participate in the interplay with polysaccharides. Several α-amylases are described to possess this type of binding sites (Gibson & Svensson 1987; Larson et al. 1994; Kadziola et al. 1998; Dauter et al. 1999; Brzozowski et al. 2000; Ramasubbu et al. 2003; Robert et al. 2003, 2005; Lyhne-Iversen et al. 2006; Vujicic-Žagar & Dijkstra, 2006; Ragunath et al. 2008). In barley α-amylase isozyme 1 (AMY1), the crystal structure of the inactive catalytic nucleophile D180A mutant in complex with maltoheptaose (Fig. 6) highlighted oligosaccharide binding at two external surface sites and the active site, respectively (Robert et al. 2005). One surface site, called "the pair of sugar tongs", was situated on the non-catalytic C-terminal domain, while the other was found on the side of the catalytic $(\beta/\alpha)_8$-barrel at a certain distance from the active site (Robert et al. 2003, 2005). The chain direction of the bound oligosaccharides in the D180A AMY1/maltoheptaose complex was such that the three molecules could not be visualized to all belong to the same polysaccharide molecule (Robert et al. 2005).

The first GH13 surface site ever reported was from barley isozyme AMY2 and identified by differential chemical modification of tryptophanyl residues using β-cyclodextrin (β-CD) for protection (Gibson & Svensson 1987). Subsequently mutagenesis in AMY1 (Søgaard et al. 1993) and crystallography of AMY2 (Kadziola et al. 1998) confirmed the carbohydrate-binding ability of this site, which contains Trp^{276} and Trp^{277} (correspond to Trp^{278} and Trp^{279} in AMY1). Very recently, this site was shown to play a dominating role for adsorption of AMY1 onto starch granules (Nielsen et al., manuscript in preparation). Also in human salivary α-amylase four

Table 1. Carbohydrate binding and enzymatic properties of "sugar tongs" mutants.

Enzyme	β-CD	Starch granules	Cl-pNPG$_7$			Amylose DP 440			Insoluble Blue Starch
	K_d mM	K_d mg mL^{-1}	k_{cat} s^{-1}	K_m mM	k_{cat}/K_m s^{-1} mM^{-1}	k_{cat} s^{-1}	K_m mg mL^{-1}	k_{cat}/K_m s^{-1} mg^{-1} mL	U mg^{-1}
Y380A AMY1[a]	1.4	5.9	19	0.669	28.4	95	0.363	261	1400
Y380M AMY1[a]	1.39	n.d.	34	0.871	39	149	0.351	424	2000
S378P AMY1[a]	0.25	0.57	59	0.861	68.5	163	0.203	802	2695
Wild-type AMY1[a]	0.2	0.47	122	1.1	111	185	0.190	973	2900
AMY2	0.24 (0.63[a])	3.5 (1.27[a])	126[a]	2.6[a]	48.5[a]	531	1.19	447	5000
M6	0.24	3.2	113[b]	2.44[b]	46.3[b]	591	1.23	484	4925
P376S M6	0.22	2.1	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	4600

[a] Bozonnet et al. (2007). [b] Fukuda et al. (2005); M6 = A42P AMY2. n.d.: not determined.

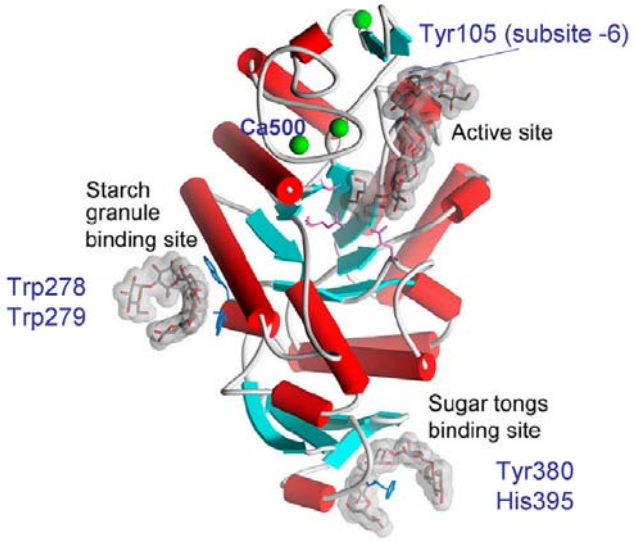


Fig. 6. Surface binding sites in barley α-amylase 1 (AMY1). The D180A inactive catalytic nucleophile mutant in complex with maltoheptaose (Robert et al. 2005).

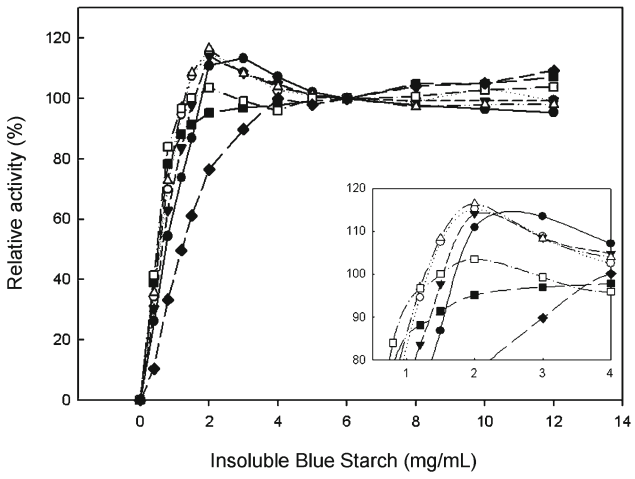


Fig. 7. Hydrolysis of insoluble Blue Starch by AMY1 mutants, AMY1 wild-type, and AMY2 wild-type. ● AMY1, ○ S378T, ▼ S378P, △ Y380F, ■ Y380A, □ Y380M, ◆ AMY2. The activity at 6.25 mg/mL insoluble Blue Starch (used in the routine assay) was normalized to 100%.

different secondary sites that bind glucose contain tryptophan (Ramasubbu et al. 2003), mutation of which to alanine eliminated the ability of the enzyme to bind to starch and bacteria, but not to bind to tooth enamel surfaces (Ragunath et al. 2008).

The other secondary binding site in AMY1 ("a pair of sugar tongs") in the non-catalytic domain C has a central Tyr^{380} (Fig. 6) that by mutational analysis was demonstrated to be important for oligosaccharide binding. Thus K_D for the starch mimic β-CD was determined by surface plasmon resonance analysis to augment from 0.20 mM for wild-type to 1.4 mM for Y380A AMY1 (Table 1; Bozonnet et al. 2007). The Y380A AMY1 mutant had a 13-fold reduced affinity and about 90% reduced catalytic efficiency towards starch granules as compared to the wild-type enzyme. Moreover, the characteristic activation of AMY1 by low concentration of the substrate insoluble Blue Starch was lost for the Y380A mutant (Fig. 7). Currently, a series of alanine mutants of Trp^{278}Trp^{279} on the catalytic $(\beta/\alpha)_8$-barrel domain of AMY1 is being studied to uncover specific roles of these residues in the interaction with starch granules and poly- and oligosaccharide substrates and their synergistic effect with for the "sugar tongs" site, respectively. Dual site mutants comprising mutation of Tyr^{105} at subsite −6, which has the highest subsite affinity at the active site (Kandra et al. 2006), and Tyr^{380} (Nielsen et al. 2008) have been constructed to study the cooperation between the active site cleft and the surface site. Moreover, Tyr^{380} at the "sugar tongs" dominated over Tyr^{105} at subsite −6 for degradation of amylose by exerting a multiple attack mechanism and by permitting hydrolysis of an insoluble starch substrate (Nielsen et al. 2008).

Polysaccharide degrading enzymes can apply a characteristic processive mechanism in which the enzyme-substrate complex executes several glycosidic bond cleavages in the same substrate molecule during a single encounter. Thus AMY1 hydrolysed amylose by such a multiple attack to release on average two oligosaccharide/maltodextrin products upon the initial cleavage in the interior part of the substrate chain (Kramhøft et al. 2005). Mutation at the "sugar tongs" in AMY1 reduced this multiple cleavage to one after the initial one per encounter. The "sugar tongs" presumably plays a role as a point of fixation of substrate at a certain distance from the active site, which provides flexibility for reorganising the substrate for multiple cleavages without breaking all contacts to the enzyme (Table 2). This re-

Table 2. Degree of multiple attack (DMA) of wild-type and "sugar tongs" mutants.

Enzyme	R_t^a	R_s^a (s^{-1})	R_p^a	DMA[b] $[(R_t/R_p) - 1]$
Wild-type AMY1[c]	138	90	48	1.9
Wild-type AMY2	248	163	85	0.5
M6	269	189	80	0.4
Y380A[d]	53	25	28	1.0
Y380M[d]	90	60	30	2.0
S378P[d]	152	105	47	2.2

[a] Amylose DP400 (1 mg/mL) was used as substrate (a.m. Kramhøft et al. 2005). R_t is the total reducing power of the reaction mixture. R_p is the reducing power of the polysaccharide fraction. R_s is the reducing power of the soluble fraction and is calculated as $R_t - R_p$.
[b] Values of DMA are means calculated from the linear rates of reducing value formation in each individual experiment.
[c] Kramhøft et al. (2005).
[d] Bozonnet et al. (2007).

sult was in agreement with the suggestion that a distant polysaccharide-binding site is needed in the processive action (Kramhøft et al. 2005). Furthermore, mutation of the "sugar tongs" slightly reduced activity towards an oligosaccharide substrate, suggesting that this site also represents a previously identified secondary site that binds oligosaccharides coupled with allosteric activation of AMY1 (Oudjeriouat et al. 2003).

Remarkably, carbohydrate did not bind at the "sugar tongs" in the crystal structure of AMY2 (Kadziola et al. 1998), although the two key residues Tyr^{380} and His^{395} were conserved. The AMY2 structure gave no useful clue to the cause of this difference from AMY1. In order to follow up on this question mutational analysis in AMY2 was pursued, however, firstly the very poor expression level of this isozyme in heterologous yeast hosts (Søgaard & Svensson 1990; Juge et al. 1996) had to be overcome. Inspired by substantial expression of AMY1–AMY2 chimeras, a structural element responsible for poor expression of AMY2 was localized to the approximately first 60 amino acid residues. Random mutational combination of the 10 sequence differences from AMY1 into AMY2 in this N-terminal segment and screening for production yield resulted in identification of a single replacement A42P AMY2 (called the M6 mutant) accompanied by greatly improved yield (Fukuda et al. 2005). M6 maintained all AMY2 characteristics tested for, i. e. kinetic constants on different substrates, recognition of the proteinaceous inhibitor BASI, stability, etc. (Fukuda et al. 2005) and M6 was therefore used as a parent for the mutational analysis of structure/function relationships in AMY2. We initiated this mutational analysis by P376S M6 of the "sugar tongs" to address the postulate that Pro^{376} in AMY2, corresponding to Ser^{378} in AMY1, was rigidifying the site and hence suppressing accommodation of oligosaccharide ligands. It turned out that P376S M6 had only slightly improved binding affinity (Table 1), and that its affinity was considerably weaker than that of AMY1 (E.S. Seo et al., unpublished results). Thus although the "sugar tongs" in AMY2 has significant, but weak affinity for oligosaccharides, it did not bind oligosaccharide ligands in the crystal structure. Preliminary data, however, on M6 Tyr^{378} mutants indicated that this residue plays a role in the binding onto starch granules (E.S. Seo et al., unpublished results).

Impact of calcium ions

Only some GH-H members, including almost all α-amylases, need calcium ions for stability and activity. α-Amylase structures display one highly conserved calcium ion, which is situated near the catalytic site (Ca500 on Figure 6); often also additional calcium or other metal ions (Na^+, Zn^{+2}) are seen in the structures. AMY1 and AMY2 show different stability dependence of calcium and mutational analysis in conjunction with differential scanning calorimetry (M. Abou Hachem et al., manuscript in preparation) showed that AMY2 was more sensitive to EDTA-induced removal of calcium ions especially at lower pH values, while at higher calcium concentration and pH values both enzymes displayed similarly high thermal stabilities. Furthermore, different mutational replacement of side chains interacting with or in the near proximity of the structural calcium ions could result in either weakening or strengthening the conformational stability depending on the mutant.

Binding to starches

Some common principles may be used to describe the mechanistic action of glycoside hydrolases for degradation of recalcitrant substrates, such as cellulose, other cell wall polysaccharides, or starches (Boraston et al. 2004, 2006; Nakai et al. 2008). A variety of CBMs representing an array of polysaccharide specificities have been identified (http://www.cazy.org/). These are widely used in studies of heterogeneous catalytic degradation of insoluble substrates as well as in applications taking advantage of the CBM affinity to direct enzyme-CBM fusion proteins to the surface of the substrate in question (Juge et al. 2006). In addition to the first identified SBD (CBM20), 7 CBM families of SBDs have been reported (Boraston et al. 2004, 2006; Machovic & Janecek 2006, 2008). Moreover, certain regions of the polypeptide chain in GH31 from plants were demonstrated to confer affinity for granular starch (Nakai et al. 2008).

We have been focusing on different SBDs of CBM20 and found that a CBM20 of plant origin from a glucan, water dikinase 3 (GWD3) has the capacity to bind onto starch granules as shown after fluorophore labeling of the recombinant domain by using confocal laser scanning microscopy to monitor binding onto starch granules (Fig. 8A). The glucan, water dikinase is a plastid-targeted enzyme involved in phosphorylation of starch (Blennow et al. 2002) resulting in increased degradability of the granule. The result demonstrates that the CBM20 indeed localizes the enzyme on the

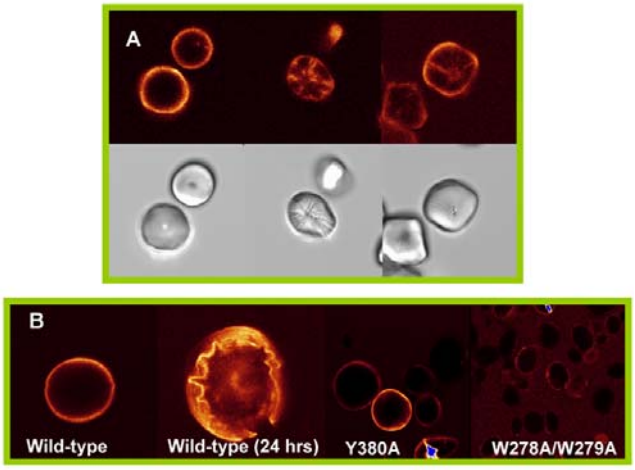


Fig. 8. Confocal laser scanning microscopy of SBD of CBM20 from GWD3 bound onto maize starch granules (A) and binding of AMY1 wild-type, Y380, and W278A/W279A onto barley starch granules (B) (courtesy C. Christiansen, M. Glaring and M.M. Nielsen).

starch molecule. Its low-millimolar affinity to starch as compared to the other members of the CBM20 family demonstrates the importance and possibility of organisms to modulate starch affinity in order to permit dynamic partitioning of enzymes to the granule surface. This domain is further characterized with respect to carbohydrate ligand affinity (C. Christiansen et al., manuscript in preparation). In GWD1 an SBD belonging to CBM45 was similarly shown to be involved in binding onto starch granules in connection with phosphorylation (Mikkelsen et al. 2006). Along the same lines the surface sites on AMY1 were implicated in binding to starch granules and confocal laser scanning microscopy similarly allowed to demonstrate loss of affinity of certain surface site mutants for starch granules (Fig. 8B).

Closing remarks

New three-dimensional structures have appeared greatly improving insight into the relationship between structure and function of starch- and related α-glucan-active enzymes. Among other aspects these structures advanced rational protein engineering of enzyme specificity. Recently, mutational analysis of sites involved in interaction with polysaccharide substrates at a distance from the active site cleft provides knowledge on how multi-site substrate interactions occur. This also includes identification of new starch-binding modules; indeed the search for well-defined α-glucan-binding modules should go on. The complete genome sequences available for a large variety of organisms have also in the area of amylolytic enzymes inspired to cloning and characterization of gene products with hitherto unreported roles. A systems biology approach may be applied to parts of the metabolism including biosynthesis of cell walls. More proteomics studies could reveal facets of synergistic action of certain enzymes along with knowledge on their appearance often in multiple forms *in vivo*. Ultimately this includes knowledge on roles of individual forms that contain post-translational modifications, an area which has received relatively little attention for the various amylases and related enzymes. Also regulatory proteins or subunits constitute an area where much is still to be investigated and which is anticipated to provide novel insight into regulation and roles in biological systems.

The modular architecture of amylolytic enzymes motivates creative design of fusion proteins with advantageous combinations of functionalities, e.g. ability to bind insoluble substrates, or manipulation of activity for various substrate categories, e.g. branched dextrins. This approach could also include more or less sophisticated engineering of off active site substrate interaction regions. Structural insight and hence understanding of the concerted action of catalytic and remote substrate subsites is highly limited. Questions remain on the mechanism of action and role of such sites in enzymatic conversions and utilization of sugars, as well as on how individual domains interact during catalysis. This has relevance for action on recalcitrant substrates. The clan GH-H contains both classical and brand-new enzymes and the collective information supplies outstanding support to advancement of fundamental knowledge on structure and function relationships as well as innovative exploitations.

Acknowledgements

The expert technical assistance of Karina Rasmussen and Sidsel Ehlers is gratefully acknowledged. We thank Mikkel Glaring for help with confocal laser scanning microscopy. The work has been supported by the EU-project CEGLYC (QLK3-CT-2001-00149), the Danish Natural Science Research Council, the Danish Research Council for Technology and Production Sciences, the Carlsberg Foundation and the Center for Advanced Food Studies (LMC). ESS held a Korea Research Foundation Grant funded by the Korean Government (MOEHRD) (KRF-2005-214-D00275) and a H.C. Ørsted postdoctoral fellowship from DTU. CC held a Ph.D. stipend from the FOBI graduate school.

References

Abbott D.W., van Bueren A.L. & Boraston A. 2007. Structural insights into the recognition of α-glucans by carbohydrate-binding modules, p. 39. In: Janecek S. (ed.), 3rd Symposium on the Alpha-Amylase Family, Programme and Abstracts, Sep 23–27, 2007, Smolenice Castle, Slovakia, Asco Arts & Science, Bratislava.

Bak-Jensen K.S., André G., Gottschalk T.E., Paës G., Tran V. & Svensson B. 2004. Tyrosine 105 and threonine 212 at outermost substrate binding subsites -6 and +4 control substrate specificity, oligosaccharide cleavage patterns, and multiple binding modes of barley α-amylase 1. J. Biol. Chem. **279:** 10093–10102.

Bak-Jensen K.S., Laugesen S., Østergaard O., Finnie C., Roepstorff P. & Svensson B. 2007. Spatio-temporal profiling and degradation of α-amylase isozymes during barley seed germination. FEBS J. **274:** 2552–2565.

Beier L., Svendsen A., Andersen C., Frandsen T.P., Borchert T.V. & Cherry J.R. 2000. Conversion of the maltogenic α-amylase Novamyl into a CGTase. Protein Eng. **13:** 509–513.

Blennow A., Engelsen S.B., Nielsen T.H., Baunsgaard L. & Mikkelsen R. 2002. Starch phosphorylation – a new front line in starch research. Trends Plant Sci.**7:** 445–450.

Boraston A.B., Bolam D.N., Gilbert H.J. & Davies G.J. 2004. Carbohydrate-binding modules: fine-tuning polysaccharide recognition. Biochem. J.**382:** 769–781.

Boraston, A.B., Healey, M., Klassen, J., Ficko-Blean, E., Lammerts van Bueren, A. & Law, V. 2006. A structural and functional analysis of α-glucan recognition by family 25 and 26 carbohydrate-binding modules reveals a conserved mode of starch recognition. J. Biol. Chem. **281:** 587–598.

Bozonnet S., Dols-Laffargue M., Fabre E., Pizzut S., Remaud-Simeon M., Monsan P. & Willemot R.M. 2002. Molecular characterization of DSR-E and an α-1,2 linkage-synthesizing dextransucrase with two catalytic domains. J. Bacteriol. **184:** 5753–5761.

Bozonnet S., Jensen M.T., Nielsen M.M., Aghajari N., Jensen M.H., Kramhøft B., Willemoës M., Tranier S., Haser R. & Svensson B. 2007. The "pair of sugar tongs" site on the non-catalytic domain C of barley α-amylase participates in substrate binding and activity. FEBS J. **274:** 5055–5067.

Brison B., Fabre E., Bozonnet S., Monsan P. & Remaud-Simeon M. 2007. How to synthesize the rare α-1,2 glucosidic linkage with an enzyme derived from a GH70 family glucansucrase?, p. 25. In: Janecek S. (ed.), 3rd Symposium on the Alpha-Amylase Family, Programme and Abstracts, Sep 23–27, 2007, Smolenice Castle, Slovakia, Asco Arts & Science, Bratislava.

Brzozowski A.M., Lawson D.M., Turkenberg J.P., Bisgård-Frantzen H., Svendsen A., Borchert T.V., Dauter Z., Wilson K.S. & Davies G.J. 2000. Structural analysis of a chimeric bacterial α-amylase. High-resolution analysis of native and ligand complexes. Biochemistry **39:** 9099–9107.

Bønsager B.C., Nielsen P.K., Abou Hachem M., Fukuda K., Prætorius-Ibba M. & Svensson B. 2005. Mutational analysis of target enzyme recognition of the β-trefoil fold barley α-amylase/subtilisin inhibitor. J. Biol. Chem. **280:** 14855–14864.

Cho M.J., Wong J.H., Marx C., Jiang W., Lemaux P.G. & Buchanan B.B. 1999. Overexpression of thioredoxin *h* leads to enhanced activity of starch debranching enzyme (pullulanase) in barley grain. Proc. Natl. Acad. Sci. USA **96:** 14641–14646.

Coutinho P.M. & Henrissat B. 1999. Carbohydrate active enzymes: an integrated database approach, pp. 3–12. In: Gilbert H.J., Davies G.J., Henrissat B. & Svensson B. (eds), Recent Advances in Carbohydrate Bioengineering, The Royal Society of Chemistry, Cambridge.

Ernst H.A., Lo Leggio L., Willemoës M., Leonard G., Blum P. & Larsen S. 2006. Structure of the *Sulfolobus solfataricus* α-glucosidase: implications for domain conservation and substrate recognition in GH31. J. Mol. Biol.**358:** 1106–1124.

Fabre E., Bozonnet S., Arcache A., Willemot R.M., Vignon M., Remaud-Simeon M. & Monsan P. 2005. Role of the two catalytic domains of DSR-E dextransucrase and their involvement in the formation of highly α-1,2 branched dextran. J. Bacteriol. **187:** 206–303.

Fukuda K., Jensen M.H., Haser R., Aghajari N. & Svensson B. 2005. Biased mutagenesis in the N-terminal region by degenerate oligonucleotide gene shuffling enhances secretory expression of barley α-amylase 2 in yeast. Protein Eng. Des. Sel. **18:** 515–526.

Giardina T., Gunning A.P., Juge N., Faulds C.B., Furniss C.S.M., Svensson B., Morris V.J. & Williamson G. 2001. Both binding sites of the starch-binding domain of *Aspergillus niger* glucoamylase are essential for inducing a conformational change in amylose. J. Mol. Biol. **313:** 1149–1159.

Gibson R.M. & Svensson B. 1987. Identification of tryptophanyl residues involved in binding of carbohydrate ligands to barley α-amylase 2. Carlsberg Res. Commun. **52:** 373–379.

Gottschalk T.E., Tull D., Aghajari N., Haser R. & Svensson B. 2001. Specificity modulation of barley α-amylase 1 by biased random mutation of a tripeptide in $\beta \rightarrow \alpha$ loop 7 of the catalytic $(\beta/\alpha)_8$-domain. Biochemistry **40:** 12844–12854.

Hondoh H., Saburi W., Mori H., Okuyama M., Nakada T., Matsuura Y. & Kimura A. 2008. Substrate recognition mechanism of α-1,6–glucosidic linkage hydrolyzing enzyme, dextran glucosidase from *Streptococcus mutans*. J. Mol. Biol. **378:** 911–920.

Janecek S., Svensson B. & MacGregor E.A. 2003. Relation between domain evolution, specificity, and taxonomy of the α-amylase family members containing a C-terminal starch-binding domain. Eur. J. Biochem. **270:** 635–645.

Janecek S., Svensson B. & MacGregor E.A. 2007. A remote but significant sequence homology between glycoside hydrolase clan GH-H and family GH31. FEBS Lett. **581:** 1261–1268.

Juge N., Andersen J.S., Tull D., Roepstorff P. & Svensson B. 1996. Overexpression, purification, and characterization of recombinant barley α-amylases 1 and 2 secreted by the methylotrophic yeast *Pichia pastoris*. Protein Express. Purif. **8:** 204–214.

Juge N., Nøhr J., Le Gal-Coëffet M.F., Kramhøft B., Furniss C.S.M., Planchot V., Archer D.B., Williamson G. & Svensson B. 2006. The activity of barley α-amylase on starch granules is enhanced by fusion of a starch binding domain from *Aspergillus niger* glucoamylase. Biochim. Biophys. Acta **1764:** 275–284.

Kadziola A., Søgaard M., Svensson B. & Haser R. 1998. Molecular structure of a barley α-amylase-inhibitor complex: implications for starch binding and catalysis. J. Mol. Biol. **279:** 205–217.

Kandra L., Abou Hachem M., Gyemant G., Kramhøft B. & Svensson B. 2006. Mapping of barley α-amylase and outer subsite mutants reveal high-affinity subsites and barriers in the long substrate binding cleft. FEBS Lett. **580:** 5049–5053.

Kim T.J., Nguyen V.D., Lee H.S., Kim M.J., Cho H.Y., Kim Y.W., Moon T.W., Park C.S., Kim J.W., Oh B.H., Lee S.B., Svensson B. & Park K.H. 2001. Modulation of the multisubstrate specificity of *Thermus* maltogenic amylase by truncation of the N-terminal domain and by a salt-induced shift of the monomer/dimer equilibrium. Biochemistry **40:** 14182–14190.

Kitamura M., Ose T., Okuyama M., Watanabe H., Yao M., Mori H., Kimura A. & Tanaka I. 2005. Crystallization and preliminary X-ray analysis of α-xylosidase from *Escherichia coli*. Acta Cryst. **F61:** 178–179.

Krajl S., Eeuwema W., Eckhardt T.H. & Dijkhuizen L. 2006. Role of asparagine 1134 in glucosidic bond and transglycosylation specificity of reuteransucrase from *Lactobacillus reuteri* 121. FEBS J. **273:** 3735–3742.

Kralj S., Van Geel-Schutten I.G., Faber E.J., van der Maarel M.J. & Dijkhuizen L. 2005. Rational transformation of *Lactobacillus reuteri* 121 reuteransucrase into a dextransucrase. Biochemistry **44:** 9206–9216.

Kramhøft B., Bak-Jensen K.S., Mori H., Juge N., Nøhr J. & Svensson B. 2005. Involvement of individual subsites and secondary substrate binding sites in multiple attack on amylose by barley α-amylase. Biochemistry **44:** 1824–1832.

Kuriki T., Kaneko H., Yanase M., Takata H., Shimada J., Handa S., Takada T., Umeyama H. & Okada S. 1996. Controlling substrate preference and transglycosylation activity of neopullulanase by manipulating steric constraint and hydrophobicity in active center. J. Biol. Chem. **271:** 17321–17329.

Lee S.S., Yu S. & Withers S.G. 2003. Detailed dissection of a new mechanism for glycoside cleavage: α-1,4–glucan lyase. Biochemistry **42:** 13081–13090.

Leemhuis H., Kragh K.M., Dijkstra B.W. & Dijkhuizen L. 2003. Engineering cyclodextrin glycosyltransferase into a starch hydrolase with a high exo-specificity. J. Biotechnol. **103:** 203–212.

Liu Y.N., Lai Y.T., Chou W I., Chang M.D. & Lyu P.C. 2007. Solution structure of family 21 carbohydrate-binding module from *Rhizopus oryzae* glucoamylase. Biochem. J. **403:** 21–30.

Lovering A.L., Lee S.S., Kim Y.W., Withers S.G. & Strynadka N.C. 2005. Mechanistic and structural analysis of a family 31 α-glycosidase and its glycosyl-enzyme intermediate. J. Biol. Chem. **280:** 2105–2115.

Lyhne-Iversen L., Hobley T.J., Kaasgaard S.G. & Harris P. 2006. Structure of *Bacillus halmapalus* α-amylase crystallized with

and without the substrate analogue acarbose and maltose. Acta Cryst. **F62:** 849–854.

MacGregor E.A. 2004. The proteinaceous inhibitor of limit dextrinase in barley and malt. Biochim. Biophys. Acta **1696:** 165–170.

MacGregor E.A., Janecek S. & Svensson B. 2001. Relationship of sequence and structure to specificity in the α-amylase family of enzymes. Biochim. Biophys. Acta **1546:** 1–20.

MacGregor E.A., Jespersen H.M. & Svensson B. 1996. A circularly permuted α-amylase-type α/β-barrel in glucan-synthesizing glucosyltransferases. FEBS Lett. **378:** 263–266.

Machovic M. & Janecek S. 2006. Starch-binding domains in the post-genome era. Cell. Mol. Life Sci. **63:** 2710–2724.

Machovic M. & Janecek S. 2008. Domain evolution in the GH13 pullulanase subfamily with focus on the carbohydrate binding module family 48. Biologia **63:** 1057–1068.

Machovic M., Svensson B., MacGregor E.A. & Janecek S. 2005. A new clan of CBM families based on bioinformatics of starch-binding domains from families CBM20 and CBM21. FEBS J. **272:** 5497–5513.

Maeda K., Finnie C., Østergaard O. & Svensson B. 2003. Identification, cloning and characterisation of two thioredoxin h isoforms, HvTrx1 and HvTrx2, from the barley seed proteome. Eur. J. Biochem. **270:** 2633–2643.

Maeda K., Finnie C. & Svensson B. 2005. Identification of thioredoxin h-reducible disulphides in proteomes by differential labeling of cysteines: Insight into recognition of proteins in barley seeds by thioredoxin h. Proteomics**5:** 1634–1644.

Maeda K., Hägglund P., Finnie C., Svensson B. & Henriksen A. 2006. Structural basis for target protein recognition by the protein disulfide reductase thioredoxin. Structure **14:** 1701–1710.

Marion C.L., Rappleye C.A., Engle J.T. & Goldman W.E. 2006. An α-(1,4)-amylase is essential for α-(1,3)-glucan production and virulence in *Histoplasma capsulatum*. Mol. Microbiol. **62:** 970–983.

Mikkelsen R., Suszkiewicz K. & Blennow A. 2006. A novel type carbohydrate-binding module identified in α-glucan, water dikinases is specific for regulated plastidial starch metabolism. Biochemistry**45**: 4674–4682.

Mori H., Bak-Jensen K.S., Gottschalk T.E., Motawia M.S., Damager I., Møller B.L. & Svensson B. 2001. Modulation of activity and substrate binding modes by single and double subsites +1/+2 and –5/–6 mutation of barley α-amylase 1. Eur. J. Biochem. **268:** 6545–6558.

Murakami T., Kanai T., Takaha H., Kuriki T. & Imanaka T. 2006. A novel branching enzyme of the GH-57 family in the hyperthermophilic archaeon *Thermococcus kodakaraensis* KOD1. J. Bacteriol. **188:** 5915–5924.

Nakai H., Tanizawa S., Ito T., Kamiya K., Kim Y.M., Yamamoto T., Matsubara K., Sakai M., Sato H., Imbe T., Okuyama M., Mori H., Chiba S., Sano Y. & Kimura A. 2008. Rice α-glucosidase isozymes and isoforms showing different starch granules-binding and -degrading ability. Biocatal. Biotransf. **26:** 104–110.

Nielsen P.K., Bønsager B.C., Berland C.R. Sigurskjold B.W. & Svensson B. 2003. Kinetics and energetics of the binding between barley α-amylase/subtilisin inhibitor and barley α-amylase 2 studied by surface plasmon resonance and isothermal titration calorimetry. Biochemistry **42:** 1478–1487.

Nordberg Karlsson E., Labes A., Turner P., Fridjohnsson O.H., Wennerberg C., Pozzo T., Hreggvidson G.O., Kristjansson J.K. & Schönheit P. 2008. Differences and similarities in enzymes from the neopullulanase subfamily isolated from thermophilic species. Biologia **63:** 1006–1014.

Okuyama M., Kaneko A., Mori H., Chiba S. & Kimura A. 2006. Structural elements to convert *Escherichia coli* α-xylosidase (YicI) into α-glucosidase. FEBS Lett. **580:** 2707–2711.

Oudjeriouat N., Moreau, Y., Santimone, M., Svensson, B., Marchis-Mouren, G. & Desseaux, V. 2003. On the mechanism of α-amylase. Eur. J. Biochem. **270:** 3871–3879.

Park J.T., Park H.S., Kang H.K., Hong J.S., Cha H., Woo E.J., Kim J.W., Kim M.J., Boos W., Lee S. & Park K.H. 2008. Oligomeric and functional properties of a debranching enzyme (TreX) from the archaeon *Sulfolobus solfataricus* P2. Biocatal. Biotransf. **26:** 76–85.

Park K.H., Kim T.J., Cheong T.K., Kim J.W., Oh B.H. & Svensson B. 2000. Structure, specificity and function of cyclomaltodextrinase, a multispecific enzyme of the α-amylase family. Biochim. Biophys. Acta **1478:** 165–185.

Pijning T., Vujičić-Žagar A., Kralj S., Eeuwema W., Dijkhuizen L. & Dijkstra B.W. 2008. Biochemical and crystallographic characterization of a glucansucrase from *Lactobacillus reuteri* 180. Biocatal. Biotransf. **26:** 12–17.

Przylas I., Terada Y., Fujii K., Takaha T., Saenger W. & Sträter N. 2000. X-ray structure of acarbose bound to amylomaltase from *Thermus aquaticus*. Eur. J. Biochem. **267:** 6903–6913.

Quezada-Calvillo R., Sim L., Ao Z, Hamaker B.R., Quaroni A., Brayer G.D., Sterchi E.E., Robayo-Torres C.C, Rose D.R. & Nichols B.L. 2008. Luminal starch substrate "brake" on maltase-glucoamylase activity is located within the glucoamylase subunit. J. Nutr. **138:** 685–692.

Ragunath C., Manuel S.G.A., Kasinathan C. & Ramasubbu N. 2008. Structure-function relationships in human salivary α-amylase: role of aromatic residues at the secondary binding sites. Biologia **63:** 1028–1034.

Ramasubbu N., Ragunath C. & Mishra P.J. 2003. Probing the role of a mobile loop in substrate binding and enzyme activity of human salivary amylase. J. Mol. Biol. **325:** 1061–1076.

Robert X., Haser R., Gottschalk T.E., Ratajczek F., Driguez H., Svensson B. & Aghajari N. 2003. The structure of barley α-amylase isozyme 1 reveals a novel role of domain C in substrate recognition and binding: "a pair of sugar tongs". Structure **11:** 973–984.

Robert X., Haser R., Mori H., Svensson B. & Aghajari N. 2005. Oligosaccharide binding to barley α-amylase. J. Biol. Chem. **280:** 32968–32978.

Rodenburg K.W., Vallée F., Juge N., Aghajari N., Guo X.J., Haser R. & Svensson B. 2000. Specific inhibition of barley α-amylase 2 by barley α-amylase/subtilisin inhibitor depends on charge interactions and can be conferred isozyme 1 by mutation. Eur. J. Biochem. **267:** 1019–1029.

Saburi W., Hondoh H., Mori H, Okuyama M. & Kimura A. 2007. Structure-function relationship and engineering of dextran glucosidase from *Streptococcus mutans*, p. 17. In: Janecek S. (ed.), 3rd Symposium on the Alpha-Amylase Family, Programme and Abstracts, Sep 23–27, 2007, Smolenice Castle, Slovakia, Asco Arts & Science, Bratislava.

Saburi W., Mori H., Saito S., Okuyama M. & Kimura A. 2006. Structural elements in dextran glucosidase responsible for high specificity to long chain substrate. Biochim. Biophys. Acta **1764:** 688–698.

Shahpiri A., Svensson B. & Finnie C. 2008. The NADPH-dependent thioredoxin reductase/thioredoxin system in germinating barley seeds: gene expression, protein profiles, and interaction between isoforms of thioredoxin *h* and thioredoxin reductase. Plant Physiol. **146:** 789–799.

Sim L., Quezada-Calvillo R., Sterchi E.E., Nichols B.L. & Rose D.R. 2008. Human intestinal maltase-glucoamylase: crystal structure of the N-terminal catalytic subunit and basis of inhibition and substrate specificity. J. Mol. Biol. **375:** 782–792.

Svensson B., Fukuda K. Nielsen P.K. & Bønsager B.C. 2004. Proteinaceous α-amylase inhibitors. Biochim. Biophys. Acta **1696:** 145–156.

Svensson B., Jespersen H., Sierks M.R. & MacGregor E.A. 1989. Sequence homology between putative raw-starch binding domains from different starch-degrading enzymes. Biochem J. **264:** 309–311.

Søgaard M., Kadziola A., Haser R. & Svensson B. 1993. Site-directed mutagenesis of histidine 93, aspartic acid 180, glutamic acid 205, histidine 290, and aspartic acid 291 at the active site and tryptophan 279 at the raw starch binding site in barley α-amylase 1. J. Biol. Chem. **268:** 22480–22484.

Søgaard M. & Svensson B. 1990. Expression of cDNAs encoding barley α-amylase 1 and 2 in yeast and characterization of the secreted proteins. Gene **94:** 173–179.

Stanley D., Rejzek M., Naested H., Dedola S., Svensson B., Field R.A., Denyer K. & Smith A.M. 2007. Probing the role of α-glucosidase (GH31) in the endosperm of germinating barley (*Hordeum vulgare*), p. 22. In: Janecek S. (ed.), 3rd Symposium on the Alpha-Amylase Family, Programme and Abstracts, Sep 23–27, 2007, Smolenice Castle, Slovakia, Asco Arts & Science, Bratislava.

Turner P., Nilsson C., Svensson D., Holst O., Gorton L. & Nordberg Karlsson E. 2005. Monomeric and dimeric cyclomaltodextrinases reveal different modes of substrate degradation. Biologia **60 (Suppl. 16):** 79–87.

Vallée F., Kadziola A., Bourne Y., Juy M., Rodenburg K.W., Svensson B. & Haser R. 1998. Crystal structure of barley α-amylase complexed with the endogenous protein inhibitor BASI at 1.9 Å resolution. Structure **6:** 649–659.

van Bueren A.L. & Boraston A.B. 2007. The structural basis of α-glucan recognition by a family 41 carbohydrate-binding module from *Thermotoga maritima*. J. Mol. Biol. **365:** 555–560.

van der Kaaij R.M., Janecek S., van der Maarel M.J.E.C. & Dijkhuizen L. 2007a. Phylogenetic and biochemical characterisation of a novel cluster of intracellular fungal α-amylase enzymes. Microbiology **153:** 4003–4015.

van der Kaaij R.M., Xuan X.L., Franken A., Ram P.J., Punt P.J., van der Maarel M.J.E.C. & Dijkhuizen L. 2007b. Characeterization of two, novel putatively cell wall associated and GPI-anchored, α-glucanotransferase enzymes of *Aspergillus niger*. Eukaryot. Cell **6:** 1178–1188.

van Leeuwen S.S., Krajl S., van Geel-Shutten I.H., Gerwig G.J., Dijkhuizen L. & Kamerling J.P. 2008. Structural analysis of the α-D-glucan (EPS180) produced by the *Lactobacillus strain* 180 glucansucrase GTF180 enzyme. Carbohydr. Res. **343:** 1237–1250.

Vujicic-Žagar A. & Dijkstra B.W. 2006. Monoclinic crystal form of *Aspergillus niger* α-amylase in complex with maltose at 1.8 Å resolution. Acta Cryst. **F62:** 716–721.

Yuan X.L., van der Kaaij R.M., van den Hondel C.A., Punt P.J., van der Maarel M.J.E.C., Dijkhuizen L. & Ram A.F. 2008. *Aspergillus niger* genome-wide analysis reveals a large number of novel α-glucan acting enzymes with unexpected expression profiles. Mol. Genet. Genomics **279:** 545–561.

Zona R., Chang-Pi-Hin F., O'Donohue M.J. & Janecek S. 2004. Bioinformatics of the glycoside hydrolase family 57 and identification of catalytic residues in amylopullulanase from *Thermococcus hydrothermalis*. Eur. J. Biochem. **271:** 2863–2872.

Received May 28, 2008
Accepted August 11, 2008

PAPER III

A CBM20 low-affinity starch binding domain from glucan, water dikinase

Camilla Christiansen[a,b], Maher Abou Hachem[b], Mikkel A. Glaring[a], Anders Viksø-Nielsen[c], Bent W. Sigurskjold[d], Birte Svensson[b,*] and Andreas Blennow[a,*]

[a] *VKR Research Centre, Pro-Active Plants, Department of Plant Biology and Biotechnology, Faculty of Life Sciences, University of Copenhagen, DK-1871 Frederiksberg, Denmark*
[b] *Enzyme and Protein Chemistry, Department of Systems Biology, Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark*
[c] *Novozymes A/S, DK-2880 Bagsvaerd, Denmark*
[d] *Department of Biology, University of Copenhagen, DK-2100 Copenhagen Ø, Denmark*

[*]Corresponding authors: Birte Svensson, E-mail: bis@bio.dtu.dk; Andreas Blennow, E-mail: abl@life.ku.dk

Key words: Bioimaging; CBM20; Glucan, water dikinase; Starch binding domain; Surface plasmon resonance

Abbreviations: CBM, carbohydrate binding module; CLSM, confocal laser scanning microscopy; GA, glucoamylase; GWD, glucan, water dikinase; SBD, starch binding domain; SPR, surface plasmon resonance; YFP, yellow fluorescent protein

Abstract

The family 20 carbohydrate binding module (CBM20) of the *Arabidopsis* starch phosphorylator glucan, water dikinase 3 (GWD3) was heterologously produced and its properties were compared to the CBM20 from a fungal glucoamylase (GA). The GWD3 CBM20 has 50-fold lower affinity for cyclodextrins than that from GA. Homology modelling identified possible structural elements responsible for this weak binding of the intracellular CBM20. Differential binding of fluorescein-labelled GWD3 and GA modules to starch granules *in vitro* was demonstrated by confocal laser scanning microscopy and yellow fluorescent protein-tagged GWD3 CBM20 expressed in tobacco confirmed binding to starch granules *in planta*.

1. Introduction

Enzymes binding reversibly to leaf starch granules are required during their synthesis or degradation as directed by photosynthesis. One recent example is the diurnal partitioning of the starch phosphorylator glucan, water dikinase 1 (GWD1) between the plastid stroma and the starch granule surface [1], highlighting redox-based regulation of starch metabolism during biosynthesis and degradation. The GWDs contain dedicated starch binding domains (SBDs) (Fig. 1). Sequence-based classification gives SBDs in carbohydrate binding module (CBM) families; 20, 21, 25, 26, 34, 41, 45, 48 and 53 [2] (http://www.cazy.org/fam/acc_CBM.html). CBM20 is encountered in archaea, bacteria, and eukaryota and occur mainly in α-amylases, β-amylases, cyclodextrin glucanosyltransferases, glucoamylases (GAs) and glucan, water dikinase (GWD).

SBDs having 90-130 amino acid residues are situated N- or C-terminally [3] and typically retain functionality in isolated form [4]. The structural fold of SBDs is conserved, consisting of a distorted *β*-barrel with 7−8 anti-parallel *β*-strands arranged in two *β*-sheets [5]. Most CBM20s possess two carbohydrate-binding sites providing bivalent interaction involving two or three conserved solvent accessible aromatic residues [6, 7]. Site 1 is shallower and more solvent exposed than site 2 that undergoes significant structural changes upon binding [6]. CBM20s bind maltoheptaose and cyclodextrins [4, 6] but their proposed main function is to attach to granular starch, thereby locally increasing the enzyme concentration at the substrate surface [8]. Another proposed role is to "unwind" α-glucan helices on the granule surface [9] resulting in a higher hydrolytic rate [10].

SBDs typically belong to extracellular amylolytic enzymes providing starch binding [7]. However, newly discovered families CBM45 [11] and CBM53 [12] in plant GWD and starch synthase III, respectively, have weak affinities for starch, suggesting that certain enzymes involved in intracellular starch metabolism in plants possess SBDs with new function. Recently, a putative CBM20 SBD was identified in partly starch granule bound GWD3 from *A. thaliana* [13, 14]. The properties of the corresponding recombinant GWD3-SBD are reported here and compared with CBM20 from *Aspergillus niger* GA [4], probing oligosaccharide binding by surface plasmon resonance (SPR) and binding to starch granules *in vitro* and *in vivo* by confocal laser scanning microscopy (CLSM). This first description of an isolated plant CBM20 demonstrates distinct function from classical fungal CBM20s, supposedly characteristic of intracellular non-hydrolytic enzymes involved in starch metabolism.

2. Materials and Methods

2.1 Homology modelling and sequence alignment

A. niger GA SBD (PDB 1KUL) was used as template (26% identity, 58% similarity for residues 509-616) to generate a GWD3-SBD homology model using Modeler v6.1 [15]. GWD3-SBD residues 170-184 lacking in the template were deleted from the model. Sequences were aligned using ClustalW.

2.2 Expression and purification of GWD3-SBD

GWD3-SBD (Asp^{68}-Gly^{184}) was cloned from full-length GWD3 cDNA [13] into *Nco*I and *Xho*I sites of pET-28a(+) (Novagen, Madison, WI), using sense:

AACATGCCATGGCAGATGGATCAGGAACGAAAGTG (*Nco*I site underlined) and antisense primer: AACCGCTCGAGACCAACATCATCATCATTACCAAC (*Xho*I site underlined) and standard molecular biology procedures. *Escherichia coli* BL21(DE3) harbouring the GWD3-SBD construct was grown at 37ºC (LB medium; 50 μg/mL kanamycin) to OD_{600} ~0.5. Expression was induced by adding isopropyl β-D-thiogalactopyranoside to 250 μM, followed by incubation at 20ºC for 20 h. After cell harvest, sonication and centrifugation, the extract was passed over a 1 mL His-trap column (QIAGEN, Hamburg, Germany). The GWD3-SBD was eluted as recommended by the manufacturer, applied to a 6 mL β-cyclodextrin Sepharose column and eluted with 10 mM β-CD. β-CD traces were removed by an additional His-trap purification and desalting using 0.5 ml Zeba Desalt Spin Columns (Pierce, Rockford, IL). Final yield was ~0.7 mg/L culture. *A. niger* recombinant GA-SBD was produced as described [16].

2.3 Surface Plasmon Resonance (SPR) analysis

Surface plasmon resonance (Biacore T100; GE Healthcare, Uppsala, Sweden) sensorgrams were recorded of binding 0.01-10 mM (15 concentrations) α-, β-, γ-CD, linear maltooligosaccharides (DP 3–7) and β-CD with 2–6 phosphate groups (Sigma, Steinheim, Germany) to biotinylated protein (900-1000 RU on streptavidin-chip). K_d values were calculated by steady-state affinity fitting (BIAevaluation 1.1 software) to the response after subtracting the reference cell signal.

2.4 In vitro *imaging of SBD binding*

Fluorescein 5-EX succinimidyl ester (Invitrogen, Paisley, UK) in 4-fold molar excess was reacted with 100 µL GWD3-SBD (1 mg/mL, 50 mM Hepes-carbonate pH 8.0, 10 mM β-CD) (final 140 µL) for 1 h at 20ºC. Fluorescein-conjugated protein was His-trap purified, desalted twice and 10 µM protein was incubated with starch granules (10 mg/mL) for 45 min at 4ºC with gentle rotation. Visualisation by CLSM (TCS SP2, Leica Microsystems, Germany) [17] used fluorescein excitation at 488 nm and recorded emission at 500–550 nm with 25% laser power and the gain varied to prevent saturation of the detector and to ensure comparable intensities. The objective used was HCX PL APO 63.0x1.20 W CORR UV. Image analysis was performed by the TCS SP2 software.

2.5 Transient expression of YFP-GWD3-SBD

The region encoding residues 1-201 of GWD3, containing the chloroplast transit peptide and SBD, was amplified using uracil-containing primers (sense: GGCTTAAUATGGAGAGCATTGGCAGCCATTG, antisense: GGTTTAAUCCCACTACTCTATCATCACCAACATCA) and fused to YFP [18]. The fusion protein was transiently expressed in *Nicotiana benthamiana* and analysed *in situ* by CLSM [18].

3. Results and Discussion

3.1 Sequence analysis and homology modelling of the GWD3-SBD

The sequence of *A. thaliana* CBM20 has 26% identity/58% similarity to *A. niger* GA-SBD compared to 50% identity/69% similarity with a putative GWD3-SBD from

Oryza sativa. Tryptophans at GWD3-SBD putative binding site 1 (W102 and W142) are conserved as are W115 (putative binding site 2) and W164 (structural in GA-SBD) (Fig. 2). The GWD3-SBD model exhibited a similar fold as GA-SBD with 7 β-strands distributed in two β-sheets, and ligand-binding site 1 was readily identified (Fig. 3A). Only F86 and W115 are conserved in GWD3-SBD from site 2, containing Y527, Y556 and W563 in GA-SBD. Noticeably, GWD3-SBD has a 7 residue deletion in the third, flexible GA-SBD loop including Y556 that is strongly involved in binding at site 2 (Fig. 2). The GWD3-SBD model (Fig. 3B) illustrated this structural difference presumably affecting the binding.

3.2 Purification and characterisation of GWD3-SBD

A GWD3-SBD construct containing residues 68-184 was the most stable one among others tested (not shown). The purified module had an apparent molecular weight in SDS-PAGE consistent with the calculated 14.2 kDa (Fig. 4). Nonetheless, the stability of GWD3-SBD was lower (T_m = 44.6 ºC, supplementary data) than that of GA-SBD (T_m = 56.7 ºC) [19]. Stability was lower after re-purification for efficient removal of β-CD indicative of ligand-stabilisation. It is also possible that interactions with the catalytic GWD3 domain provide stability to SBD when present in the native enzyme.

3.3 Affinity to cyclodextrins as determined by SPR

Differential affinity of GWD3-SBD and GA-SBD towards α-, β-, and γ-CD was demonstrated using SPR (Table 1). GWD3-SBD showed strongest binding towards α-CD, and no detectable affinity for linear maltooligosaccharides (DP 3-7) or for

phosphorylated β-CD. The β-CD affinity minimum was at neutral pH. GWD3-SBD forms expressed in *Pichia pastoris* and as a glutathione S-transferase (GST) fusion in *E. coli* (not shown) gave similar K_d values for β-CD at pH 7.0, excluding artefacts due to expression host or affinity tag. The optimum affinity for β-CD of *A. niger* GA is at pH 4.5 [20] and comparison at the GWD3-SBD optimum pH 5.5 showed affinity 50-fold higher for GA- than for GWD3-SBD (Table 1) - consistent with GA UV difference spectroscopy (K_d = 14.4 μM [21]) and NMR titration (K_d = 9 μM [22]). Affinity differences between the two GA-SBD sites assessed for the mutants, W563K (site 2) and W590K (site 1) [21] are reflected in K_d of 28 μM and 6.4 μM, respectively. This agrees with K_{d1} of 2.0 and K_{d2} of 31.3 μM obtained by fitting a two binding site model to our GA-SBD SPR data (not shown).

The 50-fold higher K_d for β-CD (Table 1) clearly identifies GWD3-SBD as a low-affinity SBD compared to GA-SBD. The physiological relevance and generality is substantiated by the plant CBM45 in GWD1 [11] and CBM53 in starch synthase III [12] both having low affinity. As for GWD3 these enzymes are predominantly found in soluble form when extracted. Furthermore, GWD1 binding to starch *in planta* is stronger in the dark [1, 23] as a possible effect of the redox potential in stroma [1]. However, GWD3 does not possess the suggested CFATC redox motif found in GWD1 and might not be redox regulated. However, the slightly lower affinity found at pH 6.5 as compared to pH 9 may suggest slightly stronger binding during the day when stromal pH increases indicating differential dynamics of binding for the GWD3-SBD in response to physiological needs.

3.4 Binding of GWD3-SBD to starch granules in vitro *and* in vivo

CLSM visualised binding of fluorescein-labelled GWD3-SBD to starch granules, including normal and waxy (high amylopectin content) barley and maize endosperm starch (type A) and normal potato tuber starch (type B). Clear fluorescence was detected at the granule surface (Fig. 5) and in channels -characteristic of maize starches [24]. The sites of label in the starch granules indicate identical binding sites for the two SBDs. However, some subtle differences in binding distribution to specific starches were observed. While both SBDs interacted internally with the waxy granules, they showed more discrete surface binding to normal cereal granules. The potato starch showed least penetration. For the maize granules the label for the GA-SBD had higher contrast than the GWD3-SBD indicating a tendency for more distinct surface localisation for GA-SBD than for GWD3-SBD suggesting that GWD3-SBD more efficiently penetrates these granules. Fluorescein labelled bovine serum albumin gave no starch granule fluorescence and free dye penetrated the granules and resulted in unspecific binding (supplementary data). *In vivo*, a YFP-GWD3-SBD fusion demonstrated affinity for starch granules by transient expression in tobacco leaves (Fig. 6). YFP fluorescence was localised in the chloroplasts (red chlorophyll autofluorescence) and was essentially confined to the starch granules.

4. Conclusion

The present functional characterisation of an isolated plant CBM20 provides evidence for *in vitro* and *in vivo* binding of a group of intracellular CBM20s with substantially lower and therefore potentially more dynamic affinity for starch granules than

CBM20s from microbial hydrolases. Structural difference between GWD3- and GA-SBD in a flexible loop at binding site 2 probably plays a role in the low affinity of GWD3-SBD.

Acknowledgements

This work was supported by the Danish Natural Science Research Council, the Danish Research Council for Technology and Production Sciences (MAH, MAG), the Carlsberg Foundation, and a FOBI PhD stipend (CC).

Reference List

[1] Mikkelsen, R., Mutenda, K.E., Mant, A., Schurmann, P. and Blennow, A. (2005) α-glucan, water dikinase (GWD): a plastidic enzyme with redox-regulated and coordinated catalytic activity and binding affinity. Proc. Natl. Acad. Sci. U. S. A 102, 1785–1790.

[2] Machovic, M. and Janecek, S. (2006) Starch-binding domains in the post-genome era. Cell Mol. Life Sci. 63, 2710–2724.

[3] Svensson, B., Jespersen, H., Sierks, M.R. and MacGregor, E.A. (1989) Sequence homology between putative raw-starch binding domains from different starch-degrading enzymes. Biochem. J. 264, 309–311.

[4] Giardina, T., Gunning, A.P., Juge, N., Faulds, C.B., Furniss, C.S.M., Svensson, B., Morris, V.J. and Williamson, G. (2001) Both binding sites of the starch-binding domain of *Aspergillus niger* glucoamylase are essential for inducing a conformational change in amylose. J. Mol. Biol. 313, 1149–1159.

[5] Sorimachi, K., Jacks, A.J., Le Gal-Coëffet, M.F., Williamson, G., Archer, D.B. and Williamson, M.P. (1996) Solution structure of the granular starch binding domain of glucoamylase from *Aspergillus niger* by nuclear magnetic resonance spectroscopy. J. Mol. Biol. 259, 970–987.

[6] Sorimachi, K., Le Gal-Coëffet, M.F., Williamson, G., Archer, D.B. and Williamson, M.P. (1997) Solution structure of the granular starch binding domain of *Aspergillus niger* glucoamylase bound to β-cyclodextrin. Structure 5, 647–661.

[7] Penninga, D., van der Veen, B.A., Knegtel, R.M.A., van Hijum, S.A.F.T., Rozeboom, H.J., Kalk, K.H., Dijkstra, B.W. and Dijkhuizen, L. (1996) The raw starch binding domain of cyclodextrin glycosyltransferase from *Bacillus circulans* strain 251. J. Biol. Chem. 271, 32777–32784.

[8] Boraston, A.B., Bolam, D.N., Gilbert, H.J. and Davies, G.J. (2004) Carbohydrate-binding modules: fine-tuning polysaccharide recognition. Biochem. J. 382, 769–781.

[9] Morris, V.J., Gunning, A.P., Faulds, C.B., Williamson, G. and Svensson, B. (2005) AFM images of complexes between amylose and *Aspergillus niger* glucoamylase mutants, native, and mutant starch binding domains: A model for the action of glucoamylase. Starch-Starke 57, 1–7.

[10] Southall, S.M., Simpson, P.J., Gilbert, H.J., Williamson, G. and Williamson, M.P. (1999) The starch-binding domain from glucoamylase disrupts the structure of starch. FEBS Lett. 447, 58–60.

[11] Mikkelsen, R., Suszkiewicz, K. and Blennow, A. (2006) A novel type carbohydrate-binding module identified in α-glucan, water dikinases is specific for regulated plastidial starch metabolism. Biochemistry 45, 4674–4682.

[12] Valdez, H.A., Busi, M.V., Wayllace, N.Z., Parisi, G., Ugalde, R.A. and Gomez-Casati, D.F. (2008) Role of the N-terminal starch-binding domains in the kinetic properties of starch synthase III from *Arabidopsis thaliana*. Biochemistry 47, 3026–3032.

[13] Baunsgaard, L., Lütken, H., Mikkelsen, R., Glaring, M.A., Pham, T.T. and Blennow, A. (2005) A novel isoform of glucan, water dikinase phosphorylates pre-phosphorylated α-glucans and is involved in starch degradation in *Arabidopsis*. Plant J. 41, 595–605.

[14] Kötting, O., Pusch, K., Tiessen, A., Geigenberger, P., Steup, M. and Ritte, G. (2005) Identification of a novel enzyme required for starch metabolism in Arabidopsis leaves. The phosphoglucan, water dikinase. Plant Physiol. 137, 242–252.

[15] Sali, A. and Blundell, T.L. (1993) Comparative protein modelling by satisfaction of spatial restraints. J. Mol. Biol. 234, 779–815.

[16] Christensen, T., Woeldike, H., Boel, E., Mortensen, S.B., Hjortshøj, K., Thim, L. and Hansen, M.T. (1988) High level expression of recombinant genes in *Aspergillus oryzae*. Bio. Technol. 6, 1419–1422.

[17] Blennow, A., Hansen, M., Schulz, A., Jorgensen, K., Donald, A.M. and Sanderson, J. (2003) The molecular deposition of transgenically modified starch in the starch granule as imaged by functional microscopy. J. Struct. Biol. 143, 229–241.

[18] Glaring, M.A., Zygadlo, A., Thorneycroft, D., Schulz, A., Smith, S.M., Blennow, A. and Baunsgaard, L. (2007) An extra-plastidial α-glucan, water dikinase from *Arabidopsis* phosphorylates amylopectin in vitro and is not necessary for transient starch degradation. J. Exp. Bot. 58, 3949–3960.

[19] Christensen, T., Svensson, B. and Sigurskjold, B.W. (1999) Thermodynamics of reversible and irreversible unfolding and domain interactions of glucoamylase from *Aspergillus niger* studied

by differential scanning and isothermal titration calorimetry. Biochemistry 38, 6300–6310.

[20] Goto, M., Tanigawa, K., Kanlyakrit, W. and Hayashida, S. (1994) The mechanism of binding of glucoamylase 1 from *Aspergillus-awamori* var. *kawachi* to cyclodextrins and raw starch. Biosci. Biotechnol. Biochem. 58, 49–54.

[21] Williamson, M.P., Le Gal-Coëffet, M.F., Sorimachi, K., Furniss, C.S., Archer, D.B. and Williamson, G. (1997) Function of conserved tryptophans in the *Aspergillus niger* glucoamylase 1 starch binding domain. Biochemistry 36, 7535–7539.

[22] Le Gal-Coëffet, M.F., Jacks, A.J., Sorimachi, K., Williamson, M.P., Williamson, G. and Archer, D.B. (1995) Expression in *Aspergillus niger* of the starch-binding domain of glucoamylase. Comparison with the proteolytically produced starch-binding domain. Eur. J. Biochem. 233, 561–567.

[23] Ritte, G., Lorberth, R. and Steup, M. (2000) Reversible binding of the starch-related R1 protein to the surface of transitory starch granules. Plant J. 21, 387–391.

[24] Glaring, M.A., Koch, C.B. and Blennow, A. (2006) Genotype-specific spatial distribution of starch molecules in the starch granule: a combined CLSM and SEM approach. Biomacromolecules 7, 2310–2320.

Figure legends

Figure 1. Domain organisation of starch phosphorylating GWD homologues from *A. thaliana* and *S. tuberosum*.

Figure 2. Alignment of bacterial and eukaryotic CBM20s. Highly conserved binding site 1 and 2 tryptophans are highlighted in blue and red, respectively, other conserved residues are in yellow. Binding residues in site 1 and 2 are marked by red and green dots, respectively. β-Strands of GA-SBD and the GWD3-SBD model are boxed. GA: glucoamylase, CGT: cyclodextrin glucanosyl transferase, MGA: maltogenic α-amylase, BMY: β-amylase, GEN: genethonin-1, LAF: laforin, 4AGT: 4-α-glucanotransferase, GWD3: glucan, water dikinase 3.

Figure 3. A) Structures of GA-SBD (PDB 1KUL, blue) and *A. thaliana* GWD3-SBD model (template 1KUL, green). B) Superimposition of binding site 1 (left) and 2 (right) of β-CD (not included for clarity) GA-SBD complex (magenta), free GA-SBD (blue), and the GWD3-SBD model (green).

Figure 4. SDS-PAGE of recombinant GWD3-SBD. Lanes M: Marker (kDa), 1: GWD3-SBD after His-trap elution, 2: GWD3-SBD after β-CD-Sepharose.

Figure 5. Interaction of fluorescein-labelled GWD3-SBD (top panels) and GA-SBD (bottom panels) with different starches: A, normal barley; B, normal maize; C, waxy barley; D, waxy maize; E, potato.

Figure 6. YFP-GWD3-SBD transiently expressed with its *A. thaliana* GWD3 transit peptide in tobacco leaf mesophyll cells, showing binding at the surface of starch granules. Scale bar: 10 μm.

Figure1

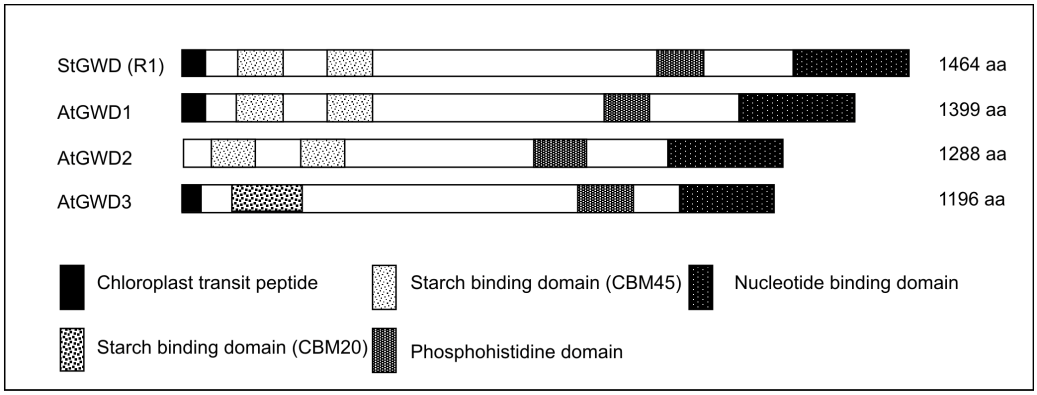

StGWD (R1)
1464 aa
AtGWD1
1399 aa
AtGWD2
1288 aa
AtGWD3
1196 aa
Chloroplast transit peptide
Starch binding domain (CBM45)
Nucleotide binding domain
Starch binding domain (CBM20)
Phosphohistidine domain

Figure 2

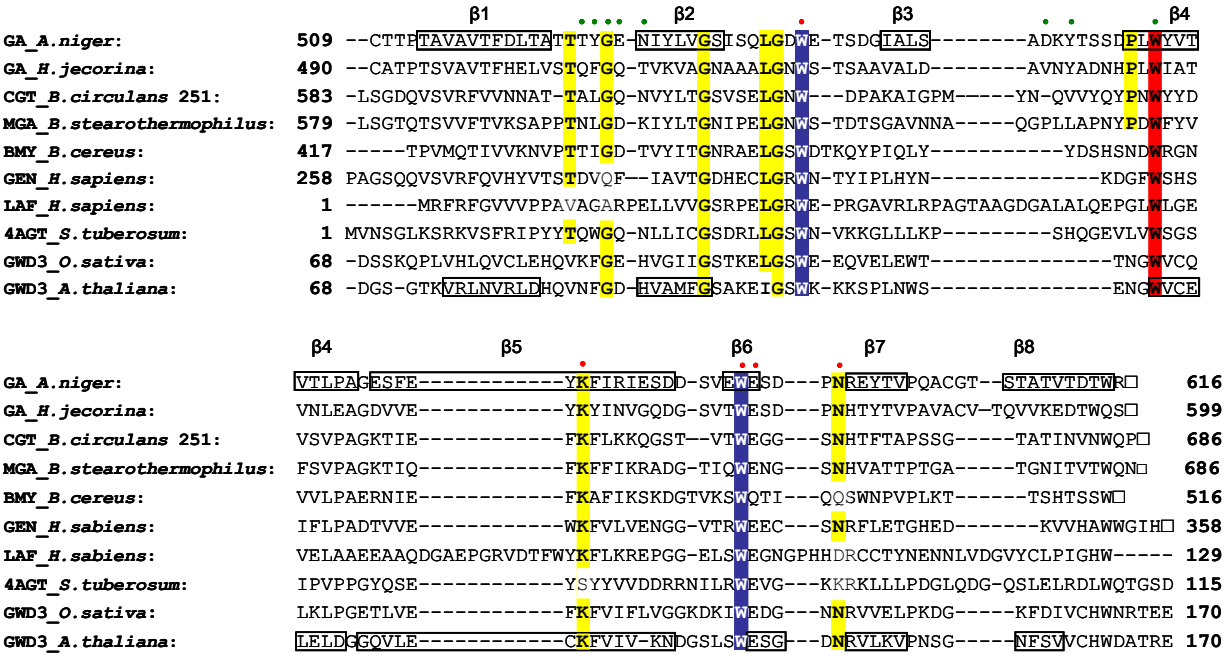

β1 β2 β3 β4
GA_A.niger: 509 --CTTPTAVAVTFDLTATTTYGE-NIYLVGSISQLGDWE-TSDGIALS--------ADKYTSSDPLWYVT
GA_H.jecorina: 490 --CATPTSVAVTFHELVSTQFGQ-TVKVAGNAAALGNWS-TSAAVALD--------AVNYADNHPLWIAT
CGT_B.circulans 251: 583 -LSGDQVSVRFVVNNAT-TALGQ-NVYLTGSVSELGNW---DPAKAIGPM-----YN-QVVYQYPNWYYD
MGA_B.stearothermophilus: 579 -LSGTQTSVVFTVKSAPPTNLGD-KIYLTGNIPELGNWS-TDTSGAVNNA-----QGPLLAPNYPDWFYV
BMY_B.cereus: 417 -----TPVMQTIVVKNVPTTIGD-TVYITGNRAELGSWDTKQYPIQLY----------YDSHSNDWRGN
GEN_H.sapiens: 258 PAGSQQVSVRFQVHYVTSTDVQF--IAVTGDHECLGRWN-TYIPLHYN----------KDGFWSHS
LAF_H.sapiens: 1 ------MRFRFGVVVPPAVAGARPELLVVGSRPELGRWE-PRGAVRLRPAGTAAGDGALALQEPGLWLGE
4AGT_S.tuberosum: 1 MVNSGLKSRKVSFRIPYYTQWGQ-NLLICGSDRLLGSWN-VKKGLLLKP---------SHQGEVLVWSGS
GWD3_O.sativa: 68 -DSSKQPLVHLQVCLEHQVKFGE-HVGIIGSTKELGSWE-EQVELEWT----------TNGWVCQ
GWD3_A.thaliana: 68 -DGS-GTKVRLNVRLDHQVNFGD-HVAMFGSAKEIGSWK-KKSPLNWS----------ENGWVCE
β4 β5 β6 β7 β8
GA_A.niger: VTLPAGESFE----------YKFIRIESDD-SVEWESD---PNREYTVPQACGT--STATVTDTWR□ 616
GA_H.jecorina: VNLEAGDVVE----------YKYINVGQDG-SVTWESD---PNHTYTVPAVACV—TQVVKEDTWQS□ 599
CGT_B.circulans 251: VSVPAGKTIE----------FKFLKKQGST—VTWEGG---SNHTFTAPSSG-----TATINVNWQP□ 686
MGA_B.stearothermophilus: FSVPAGKTIQ----------FKFFIKRADG-TIQWENG---SNHVATTPTGA-----TGNITVTWQN□ 686
BMY_B.cereus: VVLPAERNIE----------FKAFIKSKDGTVKSWQTI---QQSWNPVPLKT------TSHTSSW□ 516
GEN_H.sabiens: IFLPADTVVE----------WKFVLVENGG-VTRWEEC---SNRFLETGHED-------KVVHAWWGIH□ 358
LAF_H.sabiens: VELAAEEAAQDGAEPGRVDTFWYKFLKREPGG-ELSWEGNGPHHDRCCTYNENNLVDGVYCLPIGHW----- 129
4AGT_S.tuberosum: IPVPPGYQSE----------YSYYVVDDRRNILRWEVG---KKRKLLLPDGLQDG-QSLELRDLWQTGSD 115
GWD3_O.sativa: LKLPGETLVE----------FKFVIFLVGGKDKIWEDG---NNRVVELPKDG-----KFDIVCHWNRTEE 170
GWD3_A.thaliana: LELDGGQVLE----------CKFVIV-KNDGSLSWESG---DNRVLKVPNSG-----NFSVVCHWDATRE 170

Figure 3

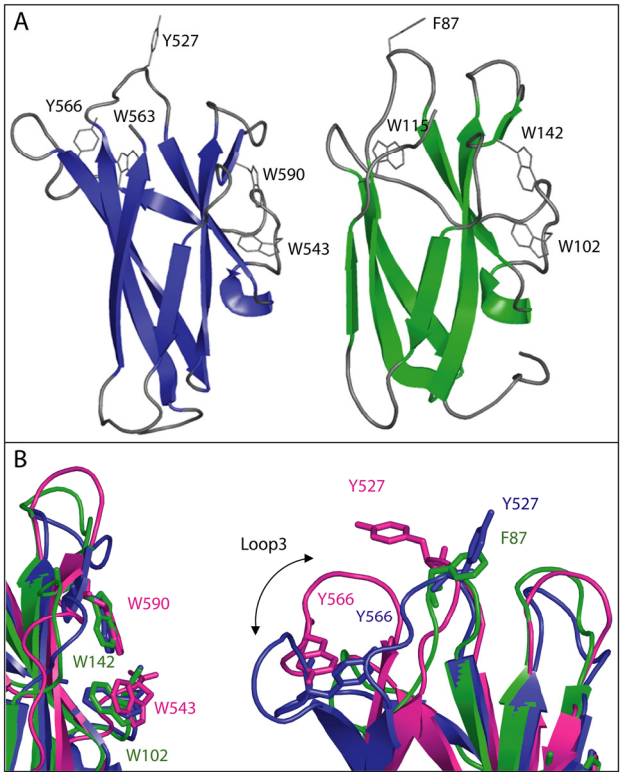

A
Y527
Y566
W563
W590
W543
F87
W115
W142
W102
B
Y527
Y527
F87
Loop3
Y566
Y566
W590
W142
W543
W102

Figure 4

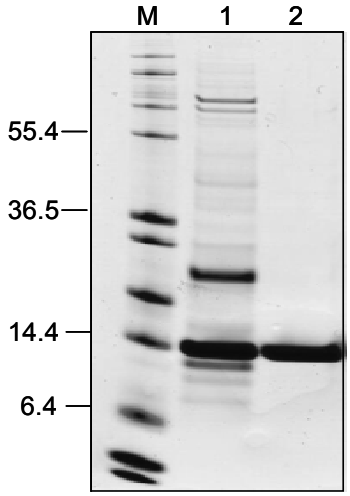


Figure 5.

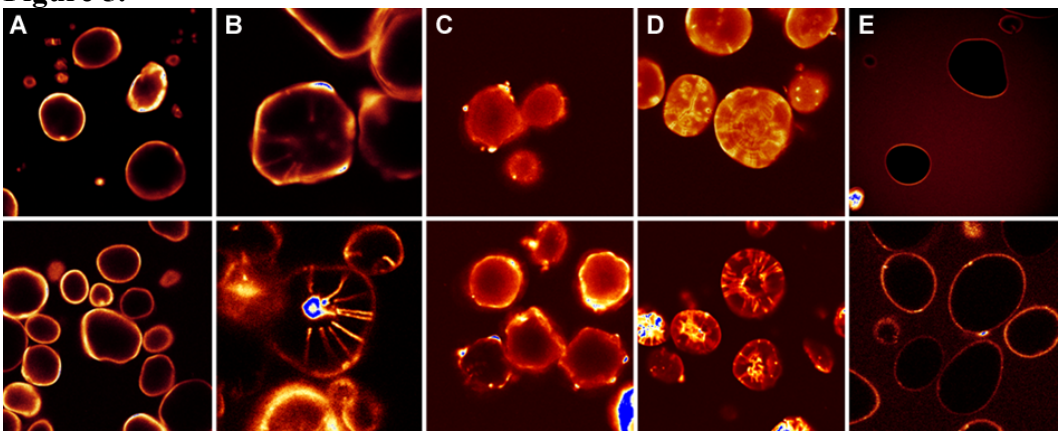


Figure 6.

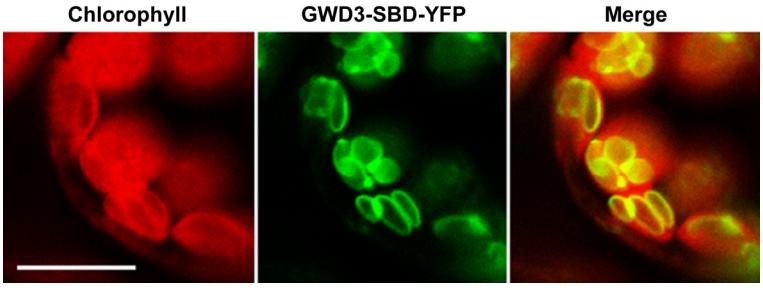


Table 1 Dissociation constants for GWD3-SBD and GA-SBD determined using SPR

CBM20	**Buffer**	**Ligand**	K_d **(mM)**
GWD3-SBD	pH 5.0	β-CD	0.62 ± 0.02
	pH 5.5	α-CD	0.22 ± 0.01
		β-CD	0.38 ± 0.07
		γ-CD	0.84 ± 0.66
	pH 6.0	β-CD	0.94 ± 0.02
	pH 6.5	β-CD	1.09 ± 0.08
	pH 7.0	α-CD	0.45 ± 0.01
		β-CD	1.07 ± 0.19
		γ-CD	1.47 ± 0.40
	pH 9.0	α-CD	0.59 ± 0.16
		β-CD	0.56 ± 0.12
		γ-CD	5.56 ± 2.07
GA-SBD	pH 5.5	β-CD	$7.5 \pm 1.30 \times 10^{-3}$

Samples were analysed in duplicates

PAPER IV

The carbohydrate binding module family 20 - diversity, structure and function

Camilla Christiansen[1,2], Maher Abou Hachem[2], Štefan Janeček[3,4], Anders Viksø-Nielsen[5], Birte Svensson[2] and Andreas Blennow[1]

[1]*VKR Research Centre, Pro-Active Plants, Department of Plant Biology and Biotechnology, Faculty of Life Sciences, University of Copenhagen, DK-1871 Frederiksberg, Denmark*
[2]*Enzyme and Protein Chemistry, Department of Systems Biology, Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark, maha@bio.dtu.dk*
[3]*Institute of Molecular Biology, Slovak Academy of Sciences, SK-84551 Bratislava, Slovakia*
[4]*Department of Biotechnology, Faculty of Natural Sciences, University of SS Cyril and Methodius, SK-91701 Trnava, Slovakia*
[5]*Novozymes A/S, DK-2880 Bagsvaerd, Denmark*

Key words: CBM20, glycoside hydrolase, SBD, starch binding domain

Abbreviations: AFM, atomic force microscopy; AMPK, adenosine monophosphate-activated protein kinase; CBM, carbohydrate binding module; CD, cyclodextrin; CGTase, cyclodextrin glucanotransferase; DP, degree of polymerisation; GA, glucoamylase; GH, glycoside hydrolase; GWD3, glucan, water dikinase 3; ITC, isothermal titration calorimetry; SEX, starch excess phenotype; SBD, starch binding domain; SPR, surface plasmon resonance

Abstract

Starch-active enzymes often possess specialised starch binding domains (SBDs) mediating attachment onto supramolecular substrates. SBDs are assigned into 9 carbohydrate binding module (CBM) families and CBM20 is the earliest and best characterised. High diversity characterises CBM20 that occurs in starch-active glycoside hydrolase (GH) families 13, 14, 15 and 77, and enzymes involved in starch or glycogen metabolism exemplified by the starch phosphorylating enzyme GWD3 from *A. thaliana* and the mammalian glycogen phosphatises, Laforins. Clear evolutionary relatedness of CBM20 to CBM21, CBM48 and CBM53 is observed suggesting a common clan hosting most known SBDs. This review surveys diversity within CBM20 and makes an evolutionary comparison to CBM families 21, 48 and 53 discussing intra and inter-family relationships. Binding and enzymatic activity data on soluble ligands and starch granules is summarised for both wild type, mutants, and chimeric fusion-proteins involving CBM20. The presence of CBM20 in amylolytic enzymes confers moderate binding affinities with dissociation constants in the low μM range on the starch mimic β-cyclodextrin., but recent findings suggest that some CBM20 members present in regulatory enzymes display significantly weaker affinities in the low mM range. Structural representatives from CBM20 including the first example of a full length glucoamylase featuring both the catalytic domain and the SBD are summarised. Architectural features and pivotal amino acid residues in binding sites 1 and 2 in CBM20 and their roles in binding are discerned. Finally, classical applications of SBDs as affinity or immobilisation tags and recent applications in the biofuel and *in planta* bioengineering are presented.

Introduction

Starch and glycogen: structure enzymatic degradability

Starch is the major energy reserve in plants and the most important energy source in human diet. The starch granule is a complex structure composed of two distinct types of glucose polymers: amylose, comprising largely unbranched α-(1→4) linked glucose residues; and amylopectin, a larger, highly branched polymer produced by the formation of α-(1→6) linkages between adjoining straight glucan chains on an α-(1→4) backbone. In the native starch granule, amylopectin and amylose molecules are organised as alternating semi-crystalline and amorphous layers that form radial growth rings (Martin and Smith, 1995). While little is known about the structure of the amorphous layers, the semi-crystalline layers are structured by short linear amylopectin chains packed together as parallel left-handed double helices. These helical segments extend from glucosidic branch-points and are further packed into concentric arrays known as the crystalline lamellae (Buléon et al., 1998). The amorphous regions of the semi-crystalline layers and the amorphous layers are composed of amylose and non-ordered amylopectin branch chains. The enzymatic degradation of starch and other insoluble polysaccharides poises a considerable challenge on glycoside hydrolases attacking them as the polysaccharide chains are often buried and poorly accessible to the active site of the enzymes. Furthermore, degradation is also governed by mass-transfer considerations in a two-phase system comprising the bulk of the medium and the surface of starch granules. Despite these common structural features, starch granules display remarkable variation depending on botanical origin and tissue or compartmentalisation (Jane et al., 1994). Such differences play a large role in the degradability of the starch granule (Buléon et al., 1998).

From the chemical point of view, starch and glycogen are very similar, but they display large differences in their molecular fine structure, physical properties, and susceptibility to enzymatic degradation. Thus, glycogen is more branched than starch and does not crystallise in water and therefore its enzymatic degradation provides a fast and short term energy reserve for rapid metabolic requirements. The chemical similarity, however, between starch and glycogen results in overlapping molecular recognition by enzymes or binding proteins targeting the two polymers.

Starch recognition and starch binding domains (SBDs)

Microbial extracellular hydrolytic enzymes that catalyze the degradation of supramolecular starch granules or plant cell wall depolymerisation typically possess a modular architecture and very often contain non-catalytic ancillary domains called carbohydrate binding modules (CBMs) that target cognate catalytic modules to specific polysaccharide structures. Binding of enzymes to insoluble polysaccharide surfaces such as starch granules or crystalline cellulose

is enhanced by CBMs that are also suggested to distort the polymers conformation and packing, thereby facilitating degradation. CBMs with affinity for starch are commonly referred to as starch binding domains (SBDs). The first discovered SBDs were assigned into CBM20 which remains the best characterised SBD family up to date. Thus, starch binding has been described for CBM20s of *e.g.* glucoamylase (GA) from *Aspergillus niger* (Sorimachi et al., 1996; Williamson et al., 1997; Sorimachi et al., 1997), cyclodextrin glucanotransferase (CGTase) from *Bacillus circulans* strain 251 (Lawson et al., 1994; Penninga et al., 1996), maltogenic α-amylase from *Geobacillus stearothermophilus* (Dauter et al., 1999), and β-amylase from *Bacillus cereus* var. *mycoides* (Mikami et al., 1999; Oyama et al., 1999).

Whereas early discovered CBM20s were from extracellular starch hydrolases secreted by fungi or bacteria (Svensson et al., 1982; Flor and Hayashida, 1983; Svensson, 1988) an increasing number of newly reported CBM20 sequences display diversity with respect to phylogenetic origin and function of joint catalytic modules. CBM20 is thus occurring in a wide spectrum of secreted or intracellular, amylolytic and non-amylolytic enzymes from plants, mammals, archaea, bacteria and fungi. Common for CBM20s is that they are appended to catalytic modules associated with starch or glycogen metabolism, and recent evidence suggests that in a few cases the enzyme activities have regulatory functions (Wang et al., 2002; Baunsgaard et al., 2005). Remarkably, the starch affinities of these regulatory proteins seem to be relatively low suggesting a much more dynamic role of their SBDs than for the counterparts present in extracellular starch degrading enzymes. The present review describes the current knowledge on structure and function of SBDs exemplified by well characterised CBM20 representatives and follows up on recent revelations regarding CBM20s with plausibly novel physiological functions different from those earlier reported for CBM20 members and SBDs from the related CBM families 21, 48, and 53.

Classification

SBDs are assigned into nine families of carbohydrate binding modules

In analogy to the glycoside hydrolase (GH) classification, CBMs are categorised in families based on amino acid sequence similarities and members of each family have a conserved structural fold (Coutinho and Henrissat, 1999) but not necessarily the same specificity. The generic term CBM thus refers to all non-catalytic carbohydrate binding domains which are grouped in 53 different families according to the CAZy classification system (see http://www.cazy.org/fam/acc_CBM.html). The first classified non-catalytic binding domains were originally defined as CBDs (cellulose-binding domains), because their specificity was towards crystalline cellulose (Gilkes et al., 1988;

Tomme et al., 1988). Similar observations were made for other polysaccharide-degrading enzymes like plant chitinases (Shinshi et al., 1990; Lerner and Raikhel, 1992) and a new term CBM, carbohydrate binding module was introduced to reflect the diverse polysaccharide specificity of such modules (Boraston et al., 1999; Abou Hachem et al., 2000). A CBM is defined as a contiguous amino acid sequence from a carbohydrate-active enzyme that folds as a separate domain and shows carbohydrate binding ability.

Currently, CBM families 20, 21, 25, 26, 34, 41, 45, 48 and 53 are reported to contain SBDs, and three-dimensional structures are available for at least one representative of each family except for CBM45 and CBM53. Despite their low sequence similarity these 7 CBM families show a very similar fold and the structural features of CBM20 will be addressed in further detail below. CBM20 has currently 276 entries displaying high phylogenetic diversity, as they are encountered in archaea (5), bacteria (143) and eukaryota (107). Five entries, all descending from industrial patents, are unclassified in the CAZy database and they are very likely of bacterial origin. Bacterial and fungal CBM20s connected with a variety of catalytic domains are most frequently found in three GH families; GH13, or the α-amylase family, GH14 containing β-amylases and GH15 containing glucoamylases (Henrissat and Davies, 1997; Janecek and Sevcik, 1999). A multitude of specificities are displayed amongst GH13 enzymes (Janecek et al., 2003) including: α-amylase (EC 3.2.1.1) (Lin et al., 2003; Lo et al., 2004); CGTase (EC 2.4.1.19) (Penninga et al., 1996); maltotetraose-forming exo-amylase (EC 3.2.1.60) (Mezaki et al., 2001); maltogenic α-amylase (EC 3.2.1.133) (Dauter et al., 1999). A few CBM20s are found in plants and include the α-glucan, water dikinase 3 (GWD3) (EC 2.7.9.4) (Baunsgaard et al., 2005) and the 4-α-glucanotransferase from GH77 (EC 2.4.1.25) (Steichen et al., 2008).

Predicted glucan binding residues

The CBM20s are typically 90-130 amino acid residues long and detailed sequence analysis revealed that there are no invariant residues in the family (Machovic et al., 2005; Machovic and Janecek, 2006a; Machovic and Janecek, 2006b). Nevertheless, the ability of CBM20 to bind to starch seems to be associated with the presence of certain consensus residues. Originally, 11 consensus residues were suggested based on multiple alignment of 8 sequences from fungi and bacteria (Svensson et al., 1989). Sequence alignment including a larger set of CBM20s, however, suggested some of these residues are more important than others due to the higher degree of conservation across the different taxa and functionalities (Janecek and Sevcik, 1999; Machovic et al., 2005). Four of the highly conserved residues are aromatic which are typically directly involved in glucan interactions.

Three-dimensional structures together with the sequence alignment approach provided support for two independent glucan binding sites on the CBM20 scaffold. Site 1 contains two conserved tryptophan residues, which in the canonical *A. niger* GA-SBD are W543 and W590 (Williamson et al., 1997). Another highly conserved residue, assigned to binding site 2, and corresponding to W563 in the GA-SBD was found to be conserved in more than 96% of the 103 CBM20 sequences analysed by (Machovic et al., 2005). The two tyrosines corresponding to Y527 and Y556 in the GA-SBD are conserved in 24 and 45 sequences, respectively, out of the analysed 103, and these two residues are also assigned to binding site 2 (Machovic et al., 2005). In addition to the original 11 consensus residues, F519 in the *A. niger* GA-SBD is 87% conserved in the 103 sequences (Machovic et al., 2005), and in remaining sequences this residue is replaced by hydrophobic residues such as isoleucine, leucine, valine. However, in CBM20 of the *Arabidopsis thaliana* starch phosphorylator GWD3 it is replaced by an arginine, whereas isoleucine or valine replace this aromatic side chain in several bacterial β-amylases (Fig 1) (Machovic et al., 2005). Finally, mutational analysis of the highly conserved W615 in GA-SBD inferred this residue a structural role as the W615K variant was difficult to produce (Williamson et al., 1997).

CBM20 is evolutionary related to CBM21, CBM48 and CBM53

Members of CBM20 were originally found exclusively at the C-terminus of various starch hydrolases and CGTases, while CBM21s were positioned N-terminally as in GA from *Rhizopus oryzae* (Svensson et al., 1989; Janecek and Sevcik, 1999). Recent bioinformatics analysis demonstrated similarity between CBM20 and CBM21 suggesting that they can constitute a new CBM clan (Machovic et al., 2005) despite their low sequence identity. Thus, CBM20 and CBM21 were predicted to have similar secondary and tertiary structures, which was recently confirmed by the first solved CBM21 structure (of *Rhizopus oryzae* GA; (Liu et al., 2007). This structure shows a conventional β-sandwich fold and an immunoglobulin-like architecture supporting the proposed clan clustering of the CBM families. Moreover, several plant 4-α-glucanotransferases, as well as GWD3 mentioned above, and also the majority of CBM20-containing unknown eukaryotic proteins (Machovic et al., 2005) possess N-terminal CBM20s. Thus CBM20 is predominantly, but not exclusively C-terminally positioned and its evolutionary relationship to CBM21 affords further analysis. In addition, CBM48 and CBM53 are also related to CBM20 (Machovic and Janecek, 2006a; Machovic and Janecek, 2006b) and therefore, representatives from these families will be included in the evolutionary analysis below.

SBDs occur with a variety of enzymatic activities

The alignment of 60 selected SBD sequences (Fig 1) clearly illustrates the close evolutionary relationship of these four CBM families, but subtle structural differences make unambiguous classification challenging. In the corresponding evolutionary tree (Fig 2) containing mostly the largest CBM20 family (33 sequences), 8 sequences from CBM21, 15 from CBM48 and 3 from the new family CBM53 further highlights the relationship. This entire set of SBD sequences was created in an effort to cover a wide spectrum of known CBM20 representatives, *i.e.* from microbial amylolytic enzymes of GH13, GH14 and GH15, the plant α-glucan, water dikinase (Baunsgaard et al., 2005), and the mammalian proteins laforin (Minassian et al., 2000) and genethonin-1 (Janecek, 2002). In addition, typical GH13 and GH15 representatives of CBM21 together with some regulatory subunits of protein phosphatases (Bork et al., 1998) proposed to form the common CBM20-21 clan (Machovic et al., 2005) were included as were CBM48 containing members of the GH13 pullulanase subfamily (Machovic and Janecek, 2008), the regulatory domains of the mammalian AMPK (adenosine monophosphate-activated protein kinase) (Polekhina et al., 2005) and the plant SEX4 (starch excess 4) (Niittyla et al., 2006) proteins which are known to be closely related to both CBM20 and CBM21 (Machovic and Janecek, 2006b). Finally, the three tandem repeated SBDs of starch synthase III from *A. thaliana* (Palopoli et al., 2006; Valdez et al., 2008), recently classified into the new family CBM53 were also included to assess evolutionary relationships between SBD affiliated to these four families (Table 1).

Different families of SBDs have subtle differences in binding sites

The alignment of amino acid sequences of SBDs and SBD-like motifs from CBM families 20, 21, 48, and 53 (Fig 1) clearly reveals the evolutionary relatedness especially concerning positions occupied by aromatic residues identified in starch-binding site 1 and 2 in CBM20 (Sorimachi et al., 1997). It is evident that the 11 CBM20 consensus residues (Svensson et al., 1989) are not strongly conserved in CBM21 or CBM48, and even within CBM20 as some members also lack some of these consensus residues (Fig 1). One prominent example is the acarviose transferase CBM20 from *Actinoplanes* sp. which was demonstrated to bind to a starch resin (Hemker et al., 2001), but where glycine replaces the conserved tryptophan in binding site 2 (Fig 1). This tryptophan, however, is conserved in CBM21 representatives originating from GH13 and GH15 amylases, but can be replaced in CBM20 and CBM21 by tyrosine or phenylalanine in non-amylolytic enzymes (Fig 1). In most CBM21 also the two tryptophans corresponding to CBM20 binding site 1 are conserved despite the fact that this site in the SBD of *Rhizopus oryzae* GH15 GA of CBM21 is located at a different position in the sequence due to a shift in the secondary structure (Liu et al., 2007).

Thus, subtle topological differences between the two families exist and shifting the position of one strand is necessary to overlap the modules (Liu et al., 2007). With respect to members of CBM48, one of the two tryptophans forming CBM20 binding site 1 (W543 in *A. niger* glucoamylase) lacks in maltooligosyltrehalose hydrolases, pullulanases and isoamylases of GH13. By contrast, the GH13 branching enzymes also together with CBM48 from the regulatory proteins AMPK (Polekhina et al., 2005), AKIN (Buitink et al., 2004) and SEX4 proteins (Niittyla et al., 2006) as well as the CBM53 repeats from starch synthase III are likely to have a functional binding site 1 despite the conservative substitution to tyrosine in some sequences at the position of W590 in the *A. niger* GA-SBD (Fig 1).

CBMs only partly cluster according to appened catalytic domains

The sequence features apparent in the CBMs alignment are reflected in the corresponding evolutionary tree (Fig 2). It should be pointed out that amylase-originating CBM20s cluster together with the starch/glycogen-binding CBM20 from laforin. The motifs recognised as possible "intermediates" (Machovic et al., 2005), *i.e.* those from GH13 amylopullulanases and carbohydrate esterases CE1, are between those CBM48s, which are most intimately related to the CBM20s, and the CBM21s in the tree. The three CBM53 repeats from starch synthase III are most closely related to CBM21 and are positioned on its border to the other cluster of CBM48s, more distant to CBM20s (Machovic and Janecek, 2006b). Noticeably modules classified in family CBM48 do not appear in a common cluster (Fig 2). Thus, CBM48s from the GH13 pullulanase subfamily (Machovic and Janecek, 2008) cluster together on a branch adjacent to CBM21 which is mostly encountered in glucoamylases, whereas the remaining CBM48s (from AMPK, AKIN1, SEX4, and related regulatory proteins) group closely together with CBM20s from genethonin-1, α-glucan, water dikinase 3, 4-α-glucanotransferase, and phosphodiesterase (Fig 2). This close relationship - originally shown before the family CBM48 was defined (Machovic and Janecek, 2006b) - clearly supports that CBM48 is grouped with CBM20 and CBM21 in a common clan.

CBM20 molecular structure

Conserved structural features of CBM20

Three-dimensional structures have been reported for 7 of the 9 SBD families defined so far; CBM20 (Sorimachi et al., 1996; Penninga et al., 1996); CBM21 (Liu et al., 2007); CBM25 and CBM26 (Boraston et al., 2006); CBM34 (Abe et al., 2004); CBM41 (Mikami et al., 2006); and CBM48 (Polekhina et al., 2005). No structures are available from CBM45 and CBM53. All solved SBD structures

have a very similar fold of a β-sandwich with an immunoglobulin-like topology. Ten CBM20 structures, either of isolated CBMs or as intact proteins possessing CBMs, have been determined using nuclear magnetic resonance (NMR) or X-ray crystallography (Table 2). Currently, the best characterised is the *A. niger* GA CBM20 (GA-SBD) which is used here as the main representative of CBM20. Its structure was determined by NMR in both free and in β-cyclodextrin complexed state and shows the well defined β-sandwich fold containing a total of eight β-strands distributed in two β-sheets (Fig 3) (Sorimachi et al., 1996; Sorimachi et al., 1997). One has five antiparallel β-strands, while the other is made from one parallel β-strand and an antiparallel β-strand pair (Jacks et al., 1995). This assembly makes an open sided distorted β-barrel, which has six loops of significant length, four of which are well defined. β-strand 3 is absent in CGTases (Klein and Schulz, 1991; Knegtel et al., 1995; Harata et al., 1996) and in the maltogenic α-amylase (Dauter et al., 1999).

The N- and C-termini are located at opposite ends of the longest axis of GA-SBD (Fig 3). A disulphide bond between the N-terminal cysteine and the loop connecting β-strands 7 and 8 (C509-C604) contributes to the structural stability and mutations of both cysteines to either glycine or serine resulted in destabilization expressed as a 10 °C reduction in unfolding temperature (T_m) and in the loss of 10 kJ/mol of stabilising free energy largely due to an unfavourable change in entropy (Sugimoto et al., 2007). The approximate dimensions of GA-SBD are 42 Å × 38 Å × 31 Å. One of the GA-SBD NMR structures has β-cyclodextrin (β-CD) bound as a starch mimic at both site 1 and 2 (Sorimachi et al., 1996) demonstrating the bivalent nature of this CBM20.

The architecture and dynamics of the binding sites

Binding site 1 involves residues W543, K578, W590, E591 and N595 and the indole rings of W543 and W590 form the central part of a carbohydrate binding platform. This shallow and solvent accessible binding site is characterised by a small contact area and very little structural change upon ligand binding (Sorimachi et al., 1997). By contrast site 2, defined by T526, Y527, G528, E529, N530, D554, Y556 and W563, is more extended and undergoes a large conformational rearrangement upon binding of β-CD attesting higher structural plasticity than site 1 (Sorimachi et al., 1997). The main change is observed in loop regions close to the C-terminal end (the left part of the protein in Fig 3). This part of binding site 2 comprises a flexible loop and in the GA-SBD β-CD complex, Y556 approaches D554 and K555 inducing a substantial change in the position of D560. The C_α atom of D560 moves more than 13 Å as compared to its position in free GA-SBD (Sorimachi et al., 1997). Also at this site, carbohydrate-protein contacts are dominated by van der Waal's stacking interactions primarily provided by Y527, Y556 and to a lesser extent W563. The requirement of the

aromatic side chains in ligand binding was confirmed by site directed mutagenesis and UV difference spectroscopy (Williamson et al., 1997). Structural plasticity at binding site 2 seems to be a general property in CBM20 and significant conformational changes were also observed in the loop of residues 460-465 in the β-amylase from *Bacillus cereus* (PDB code 1B9Z) and in the loop of residues 627-636 in the CGTase from *B. circulans* strain 251 (PDB code 1CDG), when these two domains were crystallised in complex with maltose (Lawson et al., 1994; Mikami et al., 1999).

In the crystal structure of *B. circulans* strain 251 CGTase two maltose molecules were bound at the surface of the CBM20, while a third bound on the catalytic domain (Lawson et al., 1994). Binding site 1 includes W616 and W662, which stack onto both glucose rings of the maltose. Direct hydrogen bonds with K651 and N667 and water-mediated hydrogen bonds with main chain carbonyl oxygen of W616 and E663 further enhance maltose binding. In binding site 2, Y633 stacks on one of the glucose rings of maltose, while the side chains of T598 and N627, and the main chain carbonyl oxygen of A599 and main chain amide nitrogen of G601 make direct hydrogen bonds to maltose. Another water-mediated hydrogen bond between maltose and the N603 side chain is observed. Binding site 2 is situated at the entry of the groove leading to the active site, indicating that its function may be a combination of starch binding and sequestering single glucan chains into the active site. Indeed, the side chain of the hydrophobic residue L600 in the *B. circulans* strain 251 CGTase, which is a part of binding site 2, points into the solvent and sterically confines the accessibility of this site to single carbohydrate chains (Fig 4). In other CBM20s, a phenylalanine, a tyrosine or other bulky residues at this position may serve a similar purpose. This architecture, where an aromatic or a bulky residue defines a binding site accessible to single carbohydrate chains is also observed in the recently determined structure of GA CBM21 from *R. oryzae* (Tung et al., 2008), and is reminiscent of the surface binding site observed on the C-terminal domain C of barley α-amylase 1 (Robert et al., 2003; Robert et al., 2005). Interestingly, barley α-amylase has also a second surface binding site, which is formed by two consecutive tryptophan residues, and this site bears close resemblance to binding site 1 in CBM20. It remains to be explored whether these architectural similarities confer similar functionalities in starch binding in these proteins.

Finally, a few other CBM20s than those discussed above, have been crystallised in complex with ligands (Table 2); *B. circulans* strain 8 CGTase with β-CD (PDB 1CGU) or maltotriose (PDB 5CGT); and *G. stearothermophilus* C599 maltogenic α-amylase with maltose (PDB 1QHP). All structures of the CBM:carbohydrate complexes highlight the importance of van der Waal's contacts with the indole groups of the two conserved tryptophans W543 and W590 (*A. niger* GA

numbering). Moreover, other polar residues such as N595 and K598 (*A. niger* GA numbering) in binding site 1 are likely to form direct or solvent mediated hydrogen bonds with larger ligands such as starch, and the conserved K598 packs also on W543 thus contributing to a more rigid conformation of the aromatic platform. Binding site 2 is structurally more diverse and shows differences in sequence as well as with respect to residues involved in hydrogen bonding to ligands. Noticably, no bound carbohydrate ligand was observed at site 2 in *G. stearothermophilus* C599 maltogenic α-amylase and in *B. cereus* β-amylase in complex with maltose (Mikami et al., 1999; Oyama et al., 2003). A new conserved carbohydrate binding site was identified on the catalytic domain only in bacterial β-amylases (Nanmori et al., 1993). Additional detailed studies are required to address the precise function of the two binding sites of CBM20 in catalysis.

Linker regions and interaction with the catalytic domain

The SBDs are connected differently to the catalytic domain in different enzymes. Most glucoamylases possess recognisable linker regions separating the SBD from the catalytic domain, while for example CGTases lack such linker sequences. The crystal structure of GA1 from *A. niger* including both a catalytic domain and a polypeptide linker connected SBD has not been determined. However, very recently the structure of GA from *Hypocrea jecorina* (Bott et al., 2008) was solved providing information on the spatial arrangement of the catalytic module relative the CBM20. The SBD of *H. jecorina* GA is quite similar to the average structure of *A. niger* GA-SBD determined by NMR and the two SBDs show an rmsd (root mean square deviation) of 1.7 Å for 99 aligned C_α atoms. In the *H. jecorina* GA crystal structure binding site 1 is located on the side opposite to the catalytic domain and the variable loop region of the SBD, while binding site 2 is situated close to the catalytic domain. This juxtaposition of the SBD relative to the catalytic domain is similar to the architecture of CGTase from *B. circulans* strain 251 discussed above (Lawson et al., 1994) and seems also valid for several full-length enzymes possessing CBM20 *e.g.* maltogenic α-amylase from *Geobacillus stearothermophilus C599*, and the CGTases from *B. sp. 1011* and *Thermoanaerobacterium thermosulfurigenes EM1* (Table 2) suggesting a conserved architecture within a phylogenetically diverse group of enzymes.

Another interesting question that emerges for glucoamylases where CBM20s are joined to catalytic modules by polypeptide linkers is the extent of interaction between the two modules. The crystal structure of the *H. jecorina* GA shows a well defined and intimate interaction between the SBD and the catalytic module, and the linker region has a well defined electron density and extends as a random coil anchored to the catalytic module by side chains interactions (Bott et al.,

2008). This report revealing a rather compact conformation of the enzyme, proposes that the position of the SBD relative to the catalytic module is important in directing the enzyme to regions where the starch granular matrix is disupted. By contrast, the low resolution structure of the intact GA1 from *A. niger* in solution was recently described by aid of small angle X-ray scattering (SAXS), and this model revealed an extended conformation where the highly *O*-glycosylated polypeptide linker separates the two domains of the enzyme (Jørgensen et al., 2008). The most striking difference between these two homologous enzymes is that the linker separating the SBD and the catalytic module of the *A. niger* enzyme is 22 amino acid residues longer than the *H. jecorina enzyme*. These two reports indicate that the modules of fungal glucoamulases display a certain variation in the organisation and the flexibility of the domains relative to each other promoting a fine tuned mode of action towards the natural substrates.

Ligan binding and role of CBM20 in enzyme catalysis

Binding site topology classification

On the basis of the topology of ligand-binding sites, CBMs can be classified into three distinct types, the A-, B- and C-type. The A-type is characterised by having a planar hydrophobic binding surface that recognizes highly crystalline polysaccharides such as cellulose and chitin (Boraston et al., 2004). The type-B binding site, in contrast, has the shape of a cleft or a groove with at least two discernable subsites accommodating a single polysaccharide chain (Szabo et al., 2001; Simpson et al., 2002; Pell et al., 2003). CBM20 has been assigned to this type together with the majority of identified CBMs. Typically, binding affinity of B-type CBMs strongly depends on ligand size. Thus, increased affinity up to maltononaose was demonstrated for GA-SBD, while its interaction was negligible for oligosaccharides of degree of polymerisation (DP) <4 (Belshaw and Williamson, 1993; Boraston et al., 2004). Type-B CBMs do not bind to planar surfaces typically found in highly crystalline polysaccharides such as cellulose. Moreover, in B-type as opposed to A-type binding direct hydrogen bonds play a key role in defining affinity and ligand specificity (Xie et al., 2001; Pell et al., 2003). Type-C CBMs, recognize termini of polysaccharides in a solvent-exposed binding pocket or a blind canyon having a preference for small sugars, binding optimally mono-, di- and tri-saccharrides.

The B-type binding thermodynamics in CBM20

According to binding site topology classification of CBMs (Boraston et al., 2004), ligand binding to type B modules is accompanied by an unfavourable change of entropy, which is compensated by a large favourable change in enthalpy that

dominates the binding free energy change. Indeed this is corroborated by the energetics of β-CD binding to the *A. niger* GA-SBD giving ΔG and ΔH of -26.7 kJ/mol and -58 kJ/mol respectively at pH 7.0 and 25 °C (Sugimoto et al., 2007) in good agreement with ΔG of -27.1 kJ/mol obtained for *A. awamori* GA1 at pH 4.5 and 25 °C (Goto et al., 1994b). The ΔG of binding to amylose was slightly less favourable (-22.8 kJ/mol) in agreement with the larger entropic penalty of binding to the conformationally more flexible amylose chains. The enthalpic and entropic contributions to these free energy changes were unfortunately not discerned in this case. It remains to be explored if the measured thermodynamic fingerprint of the GA-SBD is representative for other CBM20s or whether variations to this scheme exist within this diverse CBM family.

The effects of CBM binding on catalysis

Collected evidence suggests three primary effects of CBM binding in polysaccharide hydrolysis, namely "proximity effects", "substrate targeting" and "substrate structure disruption". For the "proximity effect", the CBM maintains the enzyme in close contact with substrate thereby resulting in an increased and more effective enzyme concentration at the substrate surface enhancing rates of degradation (Bolam et al., 1998). Enzymes from which CBMs were removed, either by heterologous production of a truncated form or by limited proteolysis of the native enzyme, show unchanged hydrolysis for soluble substrates, but highly reduced activity towards insoluble substrates (Svensson et al., 1982; Takahashi et al., 1985; Bolam et al., 1998; Boraston et al., 2003; Viksø-Nielsen et al., 2006). A 25 times slower hydrolytic rate for insoluble starch was reported for GA2 from *A. niger* that lacks the CBM20 domain compared to the full-length GA1 form (Svensson et al., 1982). The second proposed function is "substrate targeting" where CBMs were shown to have selective affinities for substrates like various soluble and non-soluble polysaccharides and crystalline and amorphous cellulose types (Gilkes et al., 1988; Tomme et al., 1988; Tomme et al., 1996; Carrard et al., 2000; Boraston et al., 2003). Finally, the "substrate structure disruption" effect requires multivalent interactions inherently found in most CMB20s. This was suggested initially based on the NMR structures of the GA-SBD:β-CD complex (Sorimachi et al., 1997), and gained some experimental support from data showing an increase in the initial hydrolysis rate observed when GA2 from *A. niger* was added to a starch suspension pre incubated with isolated GA-SBD (Southall et al., 1999). Further support to this derives from complexes generated by amylose and GA-SBD analysed by atomic force microscopy (AFM) (Giardina et al., 2001; Morris et al., 2005). These studies suggested that this CBM20 can act as to disentangle or force apart intertwined polysaccharide chains on supramolecular substrate surfaces hence increasing accessibility to enzymatic attack. Evidence supporting this effect, however, is indirect stemming from work

done either on isolated CBMs or on CBMs added with hydrolytic enzymes, and additional studies are needed to shed more light on the mechanism by which SBDs may exert α-glucan chain disruption in intact enzymes.

Quantifying substrate binding

Binding of soluble linear and cyclic maltooligosaccharides to CBM20 has been quantified using isothermal titration calorimetry (ITC) and UV difference spectroscopy (Goto et al., 1994b; Williamson et al., 1997; Giardina et al., 2001) leading to determination of the binding constants, stoichiometry, and thermodynamics. More recently, surface plasmon resonance (SPR) has been used to quantify binding of oligosaccharides to surface binding sites in barley α-amylase and in the CBM20 of GWD3 from *A. thaliana* (Bozonnet et al., 2007; Nielsen et al., 2008; Christiansen, unpublished) and depending on the affinity range and the type of ligands this technique is useful for evaluating oligosaccharide binding to SBDs.

For insoluble substrates, *e.g.* starch granules, the Langmuir adsorption theory has been used to describe SBD binding isotherms. The Langmuir isotherm (Eq.1) is mathematically analogous to the Michaelis-Menten model describing enzyme kinetics, and fitting the Langmuir equation to adsorption data gives an apparent dissociation constant (K_d) (some literature reports this as a K_{50} value) and a maximum saturation level often referred to as B_{max}. The B_{max} reflects the density of available binding sites at the binding surface.

$$B = \frac{B_{max}[S]}{K_d + [S]} \qquad \text{(Eq.1)}$$

Where B is the fraction of enzyme bound to starch granules and [S] is starch granules concentration. While data for SBD binding of soluble ligands is relatively straightforward and standardised, binding to starch granules is not standardised and depends on the model and experimental set up used. Such data are mostly useful for intrinsic comparisons and for comparable models. Hence if the starch concentration is maintained constant and the SBD concentration varied, the K_d values are given in molar protein concentration units (Williamson et al., 1992; Ye et al., 2004). Conversely if the SBD concentration is maintained constant and the starch concentration varied, K_d values are reported in units of starch suspension concentration, typically in mg/mL (Penninga et al., 1996). B_{max} values have been reported as SBD concentration per starch concentration or as a unit-less ratio reporting the maximum adsorbed proportion of SBD (Penninga et al., 1996; Paldi et al., 2003). Fitting the Hill equation (Eq.2) to adsorption data takes into account cooperativity effects and yields a cooperativity or Hill constant,

n, with values >1 or <1 indicating positive and negative cooperativity, respectively, while lack of cooperativity has *n* equal to 1 (Penninga et al., 1996).

$$B = \frac{B_{\max} \cdot [S]^n}{K_d + [S]^n} \qquad \text{(Eq.2)}$$

A third way to report affinity of binding to insoluble starch is to use linear adsorption isotherms that reflect equilibrium distribution of SBDs between the starch granules (bound) and the liquid phase (free), and the slope of this isotherm is defined as the adsorption constant (K_{ad}) (Chen et al., 1991). In conclusion, while binding constants determined for soluble monodisperse maltooligosaccharide ligands provide well defined data for model native starch mimics, CBMs affinity values as determined using adsorption assays to native starch granules are less well defined, but have provided data relevant for its interaction with native carbohydrate surfaces (Chen et al., 1991; Williamson et al., 1992; Dalima and Nikolov, 1994; Penninga et al., 1996; Paldi et al., 2003; Ye et al., 2004).

Enzymes containing CBM20

This section presents important cases of well characterised CBM20s with emphasis on the catalytic domains they are attached to. Different studies employ different experimental conditions such as temperatures, ionic strengths, pH values, detection methods, etc. making mutual comparison of the data problematic. Another factor complicating the comparisons is the variety of models fitted to binding data sets as discussed above. Our goal is to survey the most prominent findings for selected CBM20-containing enzymes and when possible, to give a comparative overview of existing data.

Glucoamylases

Glucoamylase is an exohydrolase catalyzing release of β-D-glucose by hydrolysis of α-1,4- and α-1,6-glucosidic linkages from non-reducing ends of starch and related oligo- and polysaccharides. GA belongs to GH15 and is an inverting enzyme produced by a wide range of microorganisms such as fungi, yeasts, bacteria and archaea (Coutinho and Reilly, 1997). The most thoroughly investigated GA is from *A. niger* that produces two major forms, GA1 and GA2, GA2 being generated by limited proteolysis of GA1 (Svensson et al., 1982; Svensson et al., 1983). The importance of the CBM originally comes from the different ability of these two forms to degrade granular starch. GA1 has three regions: (i) a catalytic domain spanning residues 1-470 including the *O*-glycosylated region of residues 441-470; (ii) a highly *O*-glycosylated linker with

residues 471-508 (iii); and a C-terminal CBM20 of residues 509-616 (Svensson et al., 1983). GA2 comprises the catalytic domain and terminates at positions 512 or 514 (Svensson et al., 1982; Svensson et al., 1983). Hence, GA2 does not contain the SBD, and consequently, GA1 and GA2 have very similar enzymatic properties towards soluble substrates (Stoffer et al., 1993), whereas GA2, in contrast to GA1 hydrolyses granular starch very slowly (Ueda, 1981; Svensson et al., 1982).

The first attempts to characterise CBM20 were conducted using *A. niger* GA examined either as the full length enzyme or the isolated SBD as obtained by limited proteolysis of GA1 together with the truncated GA2 (Belshaw and Williamson, 1990; Dalmia and Nikolov, 1991; Belshaw and Williamson, 1993). Additional information stemmed from heterologous SBD expression systems using *Escherichia coli* and fusion tags including β-galactosidase (Chen et al., 1991; Dalmia and Nikolov, 1994b) or glutathione *S*-transferase used for affinity purification and specific detection (Dalmia and Nikolov, 1994a). These strategies, however, gave very low yields and a high proportion of non-functional or degraded SBD most likely due to the poor ability of *E. coli* to produce the recombinant fusions with correct SBD disulphide bridge pivotal for stability. Later, the GA-SBD was successfully expressed *in E. coli* with smaller tags, such as a His-tag (Tanaka et al., 1998) and with an S-tag (the N-terminal part of ribonuclease A referred to as the S-peptide and used commercially as an affinity purification tag) (Paldi et al., 2003). The GA-SBD has also been produced and secreted by *A. niger* (Le Gal-Coëffet et al., 1995) and the full length GA has been heterologously expressed in *Saccharomyces cerevisiae* (Goto et al., 1994a).

For the earliest characterised recombinant CBM20 versions, the dissociation constants (K_d) of oligosaccharides were in the low micromolar range establishing this group of CBMs as moderate affinity binders (Belshaw and Williamson, 1991; Belshaw and Williamson, 1993; Goto et al., 1994b). The binding constants of proteolytically produced SBD and the heterologously expressed SBDs were determined and compared in more detail by UV difference spectroscopy and ITC. Using β-CD as ligand, the K_d values were found to vary in the range 1.7 – 23 μM (Table 3). For UV difference spectroscopy experiments, the obtained absorbance changes were very similar to those found using full length GA (Goto et al., 1994b). Moreover, the stoichiometry of the interaction (1.98 mol/mol) was essentially identical to that of GA (Belshaw and Williamson, 1991). This indicated that the β-CD interaction is specific for the SBD and not the catalytic domain and the β-CD binding to SBD inhibits starch binding competitively with an apparent inhibition constant of 11.0 ± 1.9 μM (Belshaw and Williamson, 1991). Analysing maltooligosaccharides of DP 2–9 demonstrated higher affinity with increasing degree of polymerisation, although no further increase was

observed from DP9 to DP20 (Belshaw and Williamson, 1993). This effect may be explained by an enhanced ability of the substrate to adopt a helical conformation reminiscent of the structures found in the native starch granule. Hence β-CD can be regarded as a starch mimic for detailed investigations of CBM interaction. The binding affinity of the GA SBD towards corn starch revealed K_d values from 2.6 to 19.6 μM (Williamson et al., 1992; Dalmia et al., 1995; Paldi et al., 2003) and towards potato starch of 3.3 μM (Paldi et al., 2003).

Mutational analysis provided evidence for involvement of the conserved tryptophans as suggested by sequence comparison, as single mutations at each site resulted in a decreased binding stoichiometry of β-CD from 2 to 1 per molar equivalent of SBD (Williamson et al., 1997; Giardina et al., 2001). The data also provided clear evidence for the presence of two independent binding sites with different affinity, and the K_d values of singly mutated SBDs towards maltoheptaose and β-CD remained in the low micromolar range (Williamson et al., 1997; Giardina et al., 2001). The mutant W590K in the lower affinity site 1 thus had a higher affinity of 6.4 μM as compared to the overall average affinity of wild type of 14.4 μM (Table 3), because the apparent affinity for the mutant was dominated by the high affinity at the intact site 2. Conversely, the W563K mutation at site 2 gave an affinity of 28 μM reflecting the lower affinity of site 1 (Table 3). The importance of W563 in binding to raw starch has also been confirmed by a reduction in the adsorption level from 70% of the full length GA to 11% of the W563G mutant (Goto et al., 1994a). Unfortunately, no detailed analyses were conducted on other residues and such studies would have been informative on the impact of either site on the enzymatic properties of the full-length enzyme. Investigations of amylose:SBD complexes using AFM (Giardina et al., 2001) demonstrated that abolishing one of the sites did not prevent binding at the other emphazising the mutual independence of the sites.

α-Amylases

α-Amylases of GH13 are endo-acting enzymes hydrolysing α-1,4-glucosidic bonds in starch (MacGregor et al., 2001) and related poly- and oligosaccharides with net retention of the anomeric configuration. They occur widely in animals, plants, fungi and bacteria, but only the microbial α-amylases possess CBM20s at their C-termini. Investigation of the SBDs present in the α-amylase from *Bacillus* sp. strain TS-23 provided information about the independent function of the CBM20. The SBD corresponding to residues N465-C607 was functionally expressed in *E. coli* (Lin et al., 2003) and a dissociation constant of 8.7 mg/mL was determined towards corn starch granules suggesting a relatively low affinity. The SBD showed no affinity towards avicel, xylan and chitin. A deletion of a 99 amino acid residues long segment from the C-terminal part of the intact enzyme

reduced the adsorption level to corn starch granules to less than 2% (Lo et al., 2002). The truncation of the C-terminal part, however, did not affect the raw starch hydrolytic activity as seen previously with the *A. niger* GA2, that lacks the SBD (Svensson et al., 1982). Substitutions in the SBD of T527, W545, W561, K576 and W588 (all are conserved and belong to the consensus residues) with either leucine or isoleucine was shown to significantly reduce the raw starch-binding capacity of the enzyme (Lo et al., 2004). The adsorption levels for single mutants were ~60% as compared to ~74% for a C-terminally truncated protein, which resembles the wild type in this study. The double and triple mutants gave a modest further reduction in adsorption level to ~40–50% and the triple mutant with the largest impact on binding was W545L/K576I/W588L having an adsorption level of ~20%. Also the α-amylase CBM20 from *Cryptococcus flavus* has been demonstrated to posses starch-binding ability (Wanderley et al., 2004; Galdino et al., 2008). Overall the CBM20s present in α-amylase are relatively less characterised.

Cyclodextrin glucanotransferases

CGTases (EC 2.4.1.19) of GH13 catalyze the formation of cyclodextrins from starch and related α-1,4-glucans *via* an intrachain transfer mechanism and they also catalyse coupling and disproportionation reactions through intermolecular transglycosylation. Most CGTases also have a minor but significant hydrolase activity (Penninga et al., 1995). The predominant cyclization reaction receives most attention as the common products, the α-, β-, and γ-cyclodextrins have industrial applications.

The role of CBM20s from CGTases in raw starch binding and in cyclization was investigated for the mutants W616A, W662A and W616A/W662A representing binding site 1 and Y633A at site 2 of the CGTase from *Bacillus circulans* strain 251 (Penninga et al., 1996). The wild-type CGTase showed high affinity towards starch granules with a K_d of 0.79 mg/mL, which was modestly decreased to 2.36 mg/mL in the W616A/W662A mutant. Also the Y633A mutation in site 2 affected binding having an increased K_d of 1.52 mg/mL. The maximum amount of bound enzyme was mostly affected in site 1 mutants (Table 4) indicating that this site is more important for efficient binding than site 2.

The SBD from CGTase of *Paenibacillus macerans* was expressed in *E. coli* as a β-galactosidase fusion protein (Dalmia et al., 1995) and binding properties were compared with an *A. awamori* GA-SBD β-galactosidase fusion protein also expressed in *E. coli*. Even though partial degradation of both fusion proteins was observed, the CBMs retained starch-binding activity and the K_d values for binding to raw corn starch were similar for both proteins (2.6 μM for GA- and 2.1 μM for CGTase-SBD). A truncated *P. macerans* CGTase without SBD produced in *E.*

coli was inactive and did not bind to either α-CD or acarbose affinity columns (Chang et al., 1998). An insertion of an extra linker of 6 amino acid residues between the catalytic domain and the CGTase-SBD resulted in a reduction to 59% of cyclization and to 54% of starch-hydrolyzing activity, compared with wild type enzyme. The role of the CGTase -SBD was further investigated by replacing it by the homologous (60% similarity) *A. awamori* GA-SBD (Chang et al., 1998). Although both domains were shown to retain their binding activities as independent domains, the replacement resulted in a dramatic decrease in catalytic activity to 0.03% compared to wild type CGTase. These results indicate that the SBD is an integral part of the CGTase structure and that the intimate interactions and native spatial alignment between the catalytic module and the SBD are more important for stability and catalytic performance than in case of the GA-SBD (Christensen et al., 1999) and the α-amylase SBD (Lo et al., 2002). This is also corroborated by the structure of the *Geobacillus stearothermophilus* CGTase as binding site 2 in its SBD displays a similar alignment relative to the catalytic module as observed with other CBMs discussed above (Fujiwara et al., 1992). Hence cooperativity between the binding sites on the SBD, and the catalytic module is conceivable and it has been verified by mutational studies in the two sites in the CGTase from *Bacillus circulans* strain 251, where mutations in the binding sites resulted in reduced Hill coefficient from 1.78 for the wildtype to 1.3 and 1.05 for the W616A of binding site 1 and Y633 of binding site 2, respectively (Penninga et al., 1996).

The CGTase CBM20 from *B. circulans* strain 251 has been expressed as single or two tandem SBDs in *E. coli* (Ji et al., 2004). Analysis of affinities by ITC gave lower K_d of 0.7 μM for the tandem SBD as compared to the 10 fold higher K_d of 7.7 μM for the single SBD towards soluble starch, indicating a synergistic effect of the tandem repeat of SBDs. Duplication as a possible strategy to enhance binding has been reported also for SBDs from families CBM25 (Sumitani et al., 2000) and CBM26 (Guillen et al., 2007). Interestingly, the recently discovered CBM45 SBDs occur only exclusively as tandem repeats joint to either plastidial α-amylases or glucan water dikinases 1 (Mikkelsen et al., 2006), and the recently classified CBM53 was found as triplicate repeat in the soluble starch synthase III from *Arabidopsis thaliana* (Valdez et al., 2008). It remains to be investigated whether such tandem occurrence observed for some of these SBDs may contribute to fine-tuning and possibly regulation of starch binding in response to physiological requirements.

β-Amylases

β-Amylases belong to GH14 and catalyze hydrolysis of α-1,4-glucosidic linkages with inversion of the stereochemistry at the anomeric carbon releasing β-D-

maltose from non-reducing ends of starch and maltooligosaccharides. These enzymes are produced by plants and certain bacteria and do not hydrolyse bonds next to branch points (Ye et al., 2004). Only bacterial β-amylases were found to have the ability to bind to and hydrolyze raw starch and possess a C-terminal CBM20 (Janecek and Sevcik, 1999). They exhibit maximum catalytic activity at neutral pH, while plant β-amylases have optimum activity at slightly acidic conditions. Furthermore, intrinsic rate constant of hydrolysis of α-1,4-glucosidic linkages (k_{int}) was >5 fold higher for *B. cereus* var. *mycoides* β-amylase than for plant enzymes, indicating a structural variation at the active site between β-amylases from the two different kingdoms (Nitta et al., 1996). In comparison to other CBM20s, only one binding site, corresponding to site 1 in GA-SBD is present in CBM20 of *B. cereus* var. *mycoides* β-amylase (Mikami et al., 1999; Yoon et al., 1999). Binding properties were investigated by measuring the ability of the wild type intact enzyme and mutants in the SBD binding site to bind to raw corn starch at pH 7.0 (Ye et al., 2004). The mutants W449F and W495F increased the K_d of the enzyme to 2.14 μM and 2.08 μM, respectively, compared to 0.094 μM for the wild-type. The double mutant W449F/W495F displayed essentially the same affinity as the single mutants (2.34 μM). The loss of binding energy for the three mutants was 7.1–7.3 kJ/mol. However, despite the 20-fold reduced affinity of the CBM20 binding site 1 mutants, the affinity was still high probably due to the conservative nature of the mutations performed (Ye et al., 2004). The affinity towards raw starch in the low micro molar range is comparable to that of CGTase (Dalmia et al., 1995; Ji et al., 2004) and GA (Williamson et al., 1997; Paldi et al., 2003).

Non-amylolytic CBM20 containing enzymes

The CBM20 from *A. thaliana* GWD3 is one of the newly discovered CBM20s of plant origin without associated amylolytic activity. GWD3 is a dikinase catalysing phosphorylation at the C3 position of glucose units in amylopectin (Baunsgaard et al., 2005; Kötting et al., 2005). This CBM20 has very recently been characterised as a low-affinity SBD with K_d towards β-CD of 0.4−1 mM depending on the pH (Christiansen, unpublished) and it is believed to have a regulatory role in starch degradation. Another CBM20-containing enzyme laforin, is a dual specificity mammalian protein phosphatase containing an N-terminal CBM20 (Wang et al., 2002; Abe et al., 2004; Chan et al., 2004). Laforin is a regulatory enzyme necessary for normal metabolism of storage glycogen (Lohi et al., 2005). Mutations in the *EPM2A* gene causes the Lafora form of epilepsy with an accumulation of atypical glycogen structures named polyglucan inclusion bodies, putatively arising from abnormal glycogen metabolism. These polyglucan inclusions are composed of glucan polymers with branching patterns reminiscent

of those of amylopectin from plants. Human and mouse laforins show affinity for both glycogen and starch (Wang et al., 2002; Ganesh et al., 2004; Chan et al., 2004). Recently, laforin was shown to dephosphorylate - and thereby activate - glycogen synthase kinase 3 at the serine residue S9 (Lohi et al., 2005). The other human CBM20 in the protein genethonin-1, (Janecek, 2002) clusters close to GWD3-SBD in the evolutionary tree (Fig 2). Genethonin-1 is a skeleton muscle protein, which may have a structural or regulatory role, but the function is still unknown (Bouju et al., 1998).

Another plant CBM20-like module, found in the *Arabidopsis* SEX4 protein, is classified as a CBM48 in the CAZy database. SEX4 is homologous to laforins and is a dual specificity chloroplastic phosphatase, earlier referred to as PTP-KIS1 (Fordham-Skelton et al., 2002) and binding to starch and glycogen has been verified for this protein (Niittyla et al., 2006; Kerk et al., 2006). The SEX phenotype and an apparent diurnal regulation at the *SEX4* transcript level suggest an important role in starch metabolism (Smith et al., 2004). The substrate for SEX4 has not been unambiguously determined, but it is capable of dephosphorylating amylopectin (Gentry et al., 2007).

The biological importance of enzymes possessing this category of CBM20s, the virtual absence of binding data (with the exception of the recent report of the *A. thaliana* GWD3 CBM20), and the lack of structural charterisation prompts additional studies to unravel the similiarites and differences that distinguish these CBMs from their counterparts present in amylolytic enzymes.

CBM20 and bioengineering -Fusion proteins containing CBM20

Utilising CBM20 binding functionality in protein engineering

Since isolated CBM20s generally maintain structural fold and display high affinity for starch, they have been explored as affinity purification tags after being fused to other target proteins. The CBM20s of *B. macerans* CGTase and *A. awamori* GA have both strong affinity for starch and retain their starch-binding ability when produced in *E. coli* as fusion proteins with β-galactosidase (Dalmia et al., 1995). The CBM20s as affinity purification tags were functional and after elution from a starch column by maltodextrin the purified fusion proteins maintained their β-galactosidase activity. Another chimeric enzyme, made from the α-amylase of *B. subtilis* X-23 and the CBM20 of CGTase from the alkalophilic *Bacillus* sp. strain A2-5a, maintained both raw-starch binding ability and α-amylase activity at the same levels as for the original proteins (Ohdan et al., 2000a). Moreover, a chimeric *B. stearothermophilus* leucine aminopeptidase II was connected with CBM20 from the α-amylase from *Bacillus* sp. strain TS-23 and the fusion protein was purified by adsorption-elution on raw starch (Hua et

al., 2005). These reports strongly support that several CBM20s are true modules, which retain the starch-binding ability and maintain the conformational integrity, even when separated from the original catalytic domains and introduced into a different molecular context.

Another important application of the CBM functionality is in enhancing the effecientcy starch degradation. This was demonstrated for CBM20 from the *A. niger* GA fused to the C-terminus of barley α-amylase 1 (AMY1). The rate of release of hydrolysis products from starch granules increased by a factor of 15 for this chimeric enzyme displaying also an improved K_d of 0.13 mg/mL as compared to a K_d of 0.67 mg/mL for the barley α-amylase (Juge et al., 2006). Such a fusion importantly guides development of efficient enzymes in "non-cook" or "cold hydrolysis" bioethanol strategies to circumvent the energy demanding starch gelatinisation step described in more detail in the next section.

CBM20 and starch bioengineering

SBDs have also been applied for starch bioengineering *in planta* to design new starch functionalities to be produced in starch crops. As proof of this concept, CGTase CBM20 from *B. circulans* fused with the reporter enzyme luciferace provided a model to anchor proteins inside starch granules during their biosynthesis in potato tubers (Ji et al., 2003). An advantage of this strategy is easy purification of the target protein based on simple sedimentation of the starch granules. The fusion protein was found localised throughout the matrix of the starch granules providing evidence for simultaneous protein and starch biosynthesis and deposition. Overexpression of SBDs in plants is not unproblematic since the biosynthetic machinery in the plastids is affected by SBDs, thus the starch granules were much smaller than those synthesised in the control plants (Ji et al., 2004). This effect was suggested to result from blocking of the starch granule surface by SBDs. At a very high expression level, SBDs were suggested to bind to amylose chains resulting in cross-linking of starch granules leading to agglomerated granules. Expression of tandem repeats of SBDs resulted in even more dramatic effects on starch deposition (Ji et al., 2004). In contrast, others have shown that expression of a tandem CGTase SBD from *Termoanaerobacterium termosulfurigenes* in *A. thaliana* increased the starch granule size, particularly when expressed in the *sex1-1* mutant having a null mutation in the GWD1 coding sequence involved in the phosphorylation of starch (Howitt et al., 2006). A better understanding of SBD binding on native and growing starch granules would provide insight paving the way for new applications.

Recent developments in industrial application of CBM-containing amylases

Industrial enzymatic starch hydrolysis has conventionally been initiated by gelatinisation of the starch granules using a combination of high temperature and pressure in the presence of heat stable bacterial α-amylases (Norman et al., 1997). A number of studies described different amylolytic enzymes effective in raw starch hydrolysis. Thus certain α-amylases, CGTases and glucoamylases have been described as effectively hydrolysing raw starch (Takao et al., 1987; Iefuji et al., 1996; Gawande et al., 1999). From the literature, it is evident that the presence of an SBD of CBM20 in these enzymes facilitates disruption of the starch structure resulting in a fast breakdown of the raw starch (Penninga et al., 1996; Southall et al., 1999; Morris et al., 2005; Viksø-Nielsen et al., 2006). The industrial use of these raw starch hydrolysing amylases has been hampered by poor temperature stability but, more importantly, also by a low degree of hydrolysis of raw starch compared to the current high temperature process (Nagasaka et al., 1998). Such low degrees of hydrolysis results in large losses of starch and therefore this approach is uneconomical for the starch industry (Nagasaka et al., 1998).

Recently a number of patents and publications described discovery and engineering of new CBM20-containing amylases for use in commercial scale production of glucose syrups and bioethanol (Fukuyama et al.,2006; Viksø-Nielsen et al., 2006; Wang et al., 2007). The discovery is reported of a new bacterial CBM20 containing α-amylase, which is very efficient in hydrolyzing raw starch (Viksø-Nielsen et al., 2006). It has, moreover, been demonstrated how the conventional high fructose corn syrup (HFCS) process can be turned into a simultaneous liquefaction and saccharification process using this α-amylase together with the *A. niger* GA-1 possing a CBM20. The development of new fungal CBM20-containing α-amylases for the so-called non-cook RSH (raw starch hydrolysis) fuel ethanol process is addressed (Fukuyama et al., 2006; Wang et al., 2007). These papers underline the importance of CBMs with respect to efficient raw starch hydrolysis of enzymes currently used in the fuel ethanol industry.

Concluding remarks and future perspectives

Almost two decades after the first reports on SBDs, these domains were assigned to the CBM20 family. Since then, major advances have been reached on the structural, biochemical, biophysical, and applicational aspects of this CBM family. The formidable functional and phylogenetic diversity among CBM20s relies on the same structural scaffold and a classical solvent accessible aromatic platform to mediate binding combined with several subtle albeit important structural variations that remain to be characterised. Hence, some CBM20s

members are true modules separated from catalytic domains by identifiable glycosylated polypeptide linkers as in the case of *A. niger* and several other fungal GAs. There is still a debate about the conformation of such enzymes in solution, with some data supporting a compact conformation and multiple interactions between the SBD and the catalytic domain as the recently solved crystal structure of the intact GA from *Hypocrea jecorina* (Bott et al., 2008) suggests, whereas the *A. niger* GA was shown to have an extended conformation in solution based on SAXS data (Jørgensen et al., 2008).

In other cases, the CBMs are characteristic of classical domains and showed intimate packing onto the catalytic modules rendering their function and structural stability much more dependant on the integral architecture. This is clearly evident from the virtual loss of activity upon swapping CBM20 of the *B. macerans* CGTase with that of *A. awamori* GA. Another intriguing point is the role of two binding sites present in most CBM20s. The different structure of these sites resulting in different affinities and their arrangement with respect to the catalytic domain, where the flexible site 2 is sterically confined to single carbohydrate chains and comprises an aligned extension of the groove containing the catalytic site, suggests that site 2 is likely to be associated with the catalytic mechanism, while the more rigid site 1 mediates docking onto substrate surfaces. The combination of these sites plausibly leads to effective targeting the enzymes to starch granule surfaces, where the starch structure is amorphous or disrupted and the α-glucan chains can be further disentangled and sequestered into the catalytic site. Higher resolution visualisation and assay methods would be of paramount importance in order to elucidate the precise role of each site in efficient degradation of insoluble starch granules.

The group of CBM20s which are joined to non-amylolytic enzymes is the least studied and recent reports suggest that these SBDs play important roles in starch metabolism. Structural characterisation and investigation of the thermodynamic fingerprint of a representative from this group of SBDs would be instrumental to advance understanding of the mode of action of these systems. Noticeably, SBDs linked to amylolytic enzymes seem to have evolved binding affinities in a relatively narrow range (0.1-20 μM), while CBM20s of the intracellular non-amylolytic enzymes exhibit substantially lower affinities and can occur in tandem presumably providing a more dynamic interaction with the granules.

The starch binding functionality and the modular nature of CBM20 members prompted the design of successful fusion proteins which lead to more efficient binding and degradation of starch. This approach holds a variety of potential applications in targeting or embedding enzymes on starch surfaces or inside starch granules during biosynthesis. Thus expression of CBM20 *in planta* for generating novel starch functionalities in crops and for directing proteins to starch permitting

easy purification represents obvious application concepts. Testing and thorough analysis of improvements as opposed to side effects on the crops and recovery of native protein function after harvest remain to be done.

The design of hydrolases for low temperature hydrolysis of raw starches for *e.g.* the bioethanol industries is an urgent and important task. CBM20 and other SBD types have the potential of increasing hydrolytic efficiency and rate.

Finally, the continuous updating of databases with new sequences has enabled a more robust analysis of the evolutionary relationships within CBM20 and in the larger context of the related families CBM21, CBM48, and the recently classified CBM53. This increase in the number of the sequences, however, clearly reveals the challenges of unambiguous family assignment and continuous nature of evolution.

Acknowledgements

This work was supported by Danish Natural Science Research Council, the Danish Research Council for Technology and Production (MAH) the Carlsberg Foundation, and a FOBI PhD scholarship (CC). SJ thanks the Slovak grant agency VEGA for the grant No. 2/0114/08 and the Ministry of Education of the Slovak Republic for the project AV-4/2023/08.

Reference List

Abe,A., Tonozuka,T., Sakano,Y., and Kamitori,S. (2004). Complex structures of *Thermoactinomyces vulgaris* R-47 α-amylase 1 with malto-oligosaccharides demonstrate the role of domain N acting as a starch-binding domain. J. Mol. Biol. *335*, 811-822.

Abou Hachem,M., Karlsson,E.N., Bartonek-Roxa,E., Raghothama,S., Simpson,P.J., Gilbert,H.J., Williamson,M.P., and Holst,O. (2000). Carbohydrate-binding modules from a thermostable *Rhodothermus marinus* xylanase: cloning, expression and binding studies. Biochem. J. *345*, 53-60.

Baunsgaard,L., Lütken,H., Mikkelsen,R., Glaring,M.A., Pham,T.T., and Blennow,A. (2005). A novel isoform of glucan, water dikinase phosphorylates pre-phosphorylated α-glucans and is involved in starch degradation in *Arabidopsis*. Plant J. *41*, 595-605.

Belshaw,N.J. and Williamson,G. (1990). Production and purification of a granular-starch-binding domain of glucoamylase 1 from *Aspergillus niger*. FEBS Lett. *269*, 350-353.

Belshaw,N.J. and Williamson,G. (1991). Interaction of β-cyclodextrin with the granular starch binding domain of glucoamylase. Biochim. Biophys. Acta *1078*, 117-120.

Belshaw,N.J. and Williamson,G. (1993). Specificity of the binding domain of glucoamylase 1. Eur. J. Biochem. *211*, 717-724.

Bolam,D.N., Ciruela,A., Queen-Mason,S., Simpson,P., Williamson,M.P., Rixon,J.E., Boraston,A., Hazlewood,G.P., and Gilbert,H.J. (1998). *Pseudomonas* cellulose-binding domains mediate their effects by increasing enzyme substrate proximity. Biochem. J. *331 (Pt 3)*, 775-781.

Boraston,A.B., Bolam,D.N., Gilbert,H.J., and Davies,G.J. (2004). Carbohydrate-binding modules: fine-tuning polysaccharide recognition. Biochem. J. *382*, 769-781.

Boraston,A.B., Healey,M., Klassen,J., Ficko-Blean,E., Lammerts van,B.A., and Law,V. (2006). A structural and functional analysis of α-glucan recognition by family 25 and 26 carbohydrate-binding modules reveals a conserved mode of starch recognition. J. Biol. Chem. *281*, 587-598.

Boraston,A.B., Kwan,E., Chiu,P., Warren,R.A., and Kilburn,D.G. (2003). Recognition and hydrolysis of noncrystalline cellulose. J. Biol. Chem. *278*, 6120-6127.

Boraston,A.B., Mclean,B.W., Kormos,J.M., Alam,M., Gilkes,N.R., Haynes,C.A., Tomme,P., Kilburn,D.G., and Warren,R.A.J. (1999). Carbohydrate-binding modules: diversity of structure and function. In Recent Advances in Carbohydrate Bioengineering (Gilbert, H.J., Davies, G.J., Henrissat, B. and Svensson, B., eds), pp. 202-211, Royal Society of Chemistry, Cambridge.

Bork,P., Dandekar,T., Eisenhaber,F., and Huynen,M. (1998). Characterization of targeting domains by sequence analysis: glycogen-binding domains in protein phosphatases. J. Mol. Med. *76*, 77-79.

Bott,R., Saldajeno,M., Cuevas,W., Ward,D., Scheffers,M., Aehle,W., Karkehabadi,S., Sandgren,M., and Hansson,H. (2008). Three-dimensional structure of an intact glycoside hydrolase family 15 glucoamylase from *Hypocrea jecorina*. Biochemistry *47*, 5746-5754.

Bouju,S., Lignon,M.F., Pietu,G., Le Cunff,M., Leger,J.J., Auffray,C., and Dechesne,C.A. (1998). Molecular cloning and functional expression of a novel human gene encoding two 41-43 kDa skeletal muscle internal membrane proteins. Biochem. J. *335*, 549-556.

Bozonnet,S., Jensen,M.T., Nielsen,M.M., Aghajari,N., Jensen,M.H., Kramhoft,B., Willemoes,M., Tranier,S., Haser,R., and Svensson,B. (2007). The 'pair of sugar tongs' site on the non-catalytic domain C of barley α-amylase participates in substrate binding and activity. FEBS J. *274*, 5055-5067.

Buitink,J., Thomas,M., Gissot,L., and Leprince,O. (2004). Starvation, osmotic stress and desiccation tolerance lead to expression of different genes of the regulatory β and γ subunits of the SnRK1 complex in germinating seeds of *Medicago truncatula*. Plant Cell Environ. *27*, 55-67.

Buléon,A., Colonna,P., Planchot,V., and Ball,S. (1998). Starch granules: structure and biosynthesis. Int. J. Biol. Macromol. *23*, 85-112.

Carrard,G., Koivula,A., Soderlund,H., and Beguin,P. (2000). Cellulose-binding domains promote hydrolysis of different sites on crystalline cellulose. Proc. Natl. Acad. Sci. U. S. A *97*, 10342-10347.

Chan,E.M., Ackerley,C.A., Lohi,H., Ianzano,L., Cortez,M.A., Shannon,P., Scherer,S.W., and Minassian,B.A. (2004). Laforin preferentially binds the neurotoxic starch-like polyglucosans, which form in its absence in progressive myoclonus epilepsy. Hum. Mol. Genet. *13*, 1117-1129.

Chang,H.Y., Irwin,P.M., and Nikolov,Z.L. (1998). Effects of mutations in the starch-binding domain of *Bacillus macerans* cyclodextrin glycosyltransferase. J. Biotechnol. *65*, 191-202.

Chen,L.J., Ford,C., and Nikolov,Z. (1991). Adsorption to starch of a β-galactosidase fusion protein containing the starch-binding region of *Aspergillus* glucoamylase. Gene *99*, 121-126.

Christensen,T., Svensson,B., and Sigurskjold,B.W. (1999). Thermodynamics of reversible and irreversible unfolding and domain interactions of glucoamylase from *Aspergillus niger* studied by differential scanning and isothermal titration calorimetry. Biochemistry *38*, 6300-6310.

Christiansen,C., Abou Hachem,M., Glaring,M.A., Viksø-Nielsen,A., Sigurskjold,B.W., Svensson,B., and Blennow,A. (2009). A CBM20 low-affinity starch binding domain from Glucan, Water Dikinase. (submitted).

Coutinho,P.M. and Henrissat,B. (1999). Carbohydrate-active enzymes: An integrated database approach. In Recent Advances in Carbohydrate Bioengineering (Gilbert, H.J., Davies, G.J., Henrissat, B. and Svensson, B., eds), pp. 3-12, Royal Society of Chemistry, Cambridge.

Coutinho,P.M. and Reilly,P.J. (1997). Glucoamylase structural, functional, and evolutionary relationships. Proteins *29*, 334-347.

Dalmia,B.K. and Nikolov,Z.L. (1994). Characterization of a ß-galactosidase fusion protein containing the starch-binding domain of *Aspergillus* glucoamylase. Enzyme Microb. Technol. *16*, 18-23.

Dalmia,B.K. and Nikolov,Z.L. (1991). Characterization of glucoamylase adsorption to raw starch. Enzyme Microb. Technol. *13*, 982-990.

Dalmia,B.K. and Nikolov,Z.L. (1994a). A glutathione S-transferase fusion protein with the starch-binding domain of *Aspergillus glucoamylase*. Ann. N. Y. Acad. Sci. *721*, 160-167.

Dalmia,B.K. and Nikolov,Z.L. (1994b). Characterization of a ß-galactosidase fusion protein containing the starch-binding domain of *Aspergillus* glucoamylase. Enzyme Microb. Technol. *16*, 18-23.

Dalmia,B.K., Schutte,K., and Nikolov,Z.L. (1995). Domain E of *Bacillus Macerans* Cyclodextrin Glucanotransferase: An Independent Starch-Binding Domain. Biotechnol. Bioeng. *47*, 575-584.

Dauter,Z., Dauter,M., Brzozowski,A.M., Christensen,S., Borchert,T.V., Beier,L., Wilson,K.S., and Davies,G.J. (1999). X-ray structure of Novamyl, the five-domain "maltogenic" α-amylase from *Bacillus stearothermophilus:* maltose and acarbose complexes at 1.7Å resolution. Biochemistry *38*, 8385-8392.

Flor,P.Q. and Hayashida,S. (1983). Production and characteristics of raw starch-digesting glucoamylase O from a protease-negative, glycosidase-negative *Aspergillus awamori* var. *kawachi* mutant. Appl. Environ. Microbiol. *45*, 905-912.

Fordham-Skelton,A.P., Chilley,P., Lumbreras,V., Reignoux,S., Fenton,T.R., Dahm,C.C., Pages,M., and Gatehouse,J.A. (2002). A novel higher plant protein tyrosine phosphatase interacts with SNF1-related protein kinases via a KIS (kinase interaction sequence) domain. Plant J. *29*, 705-715.

Fujiwara,S., Kakihara,H., Sakaguchi,K., and Imanaka,T. (1992). Analysis of mutations in cyclodextrin glucanotransferase from *Bacillus stearothermophilus* which affect cyclization characteristics and thermostability. J. Bacteriol. *174*, 7478-7481.

Fukuyama, S., Matsui, T., Soong, C., Allain, E., Viksø-Nielsen, A., Udagawa, H., Liu, Y., Duan, J., Wu, W. (Novozymes A/S, Denmark and Novozymes North America, INC, US) Enzymes for starch processing. Patent: WO06069290

Galdino,A.S., Ulhoa,C.J., Moraes,L.M., Prates,M.V., Bloch,C., Jr., and Torres,F.A. (2008). Cloning, molecular characterization and heterologous expression of AMY1, an α-amylase gene from *Cryptococcus flavus*. FEMS Microbiol. Lett. *280*, 189-194.

Ganesh,S., Tsurutani,N., Suzuki,T., Hoshii,Y., Ishihara,T., gado-Escueta,A.V., and Yamakawa,K. (2004). The carbohydrate-binding domain of Lafora disease protein targets Lafora polyglucosan bodies. Biochem. Biophys. Res. Commun. *313*, 1101-1109.

Gawande,B.N., Goel,A., Patkar,A.Y., and Nene,S.N. (1999). Purification and properties of a novel raw starch degrading cyclomaltodextrin glucanotransferase from *Bacillus firmus*. Appl. Microbiol. Biotechnol. *51*, 504-509.

Gentry,M.S., Dowen III,R.H., Worby,C.A., Mattoo,S., Ecker,J.R., and Dixon,J.E. (2007). The phosphatase laforin crosses evolutionary boundaries and links carbohydrate metabolism to neuronal disease. J. Cell Biol. *178*, 477-488.

Giardina,T., Gunning,A.P., Juge,N., Faulds,C.B., Furniss,C.S.M., Svensson,B., Morris,V.J., and Williamson,G. (2001). Both binding sites of the starch-binding domain of *Aspergillus niger* glucoamylase are essential for inducing a conformational change in amylose. J. Mol. Biol. *313*, 1149-1159.

Gilkes,N.R., Warren,R.A., Miller,R.C., Jr., and Kilburn,D.G. (1988). Precise excision of the cellulose binding domains from two *Cellulomonas fimi* cellulases by a homologous protease and the effect on catalysis. J. Biol. Chem. *263*, 10401-10407.

Goto,M., Semimaru,T., Furukawa,K., and Hayashida,S. (1994a). Analysis of the raw starch-binding domain by mutation of a glucoamylase from *Aspergillus awamori* var. *kawachi* expressed in *Saccharomyces cerevisiae*. Appl. Environ. Microbiol. *60*, 3926-3930.

Goto,M., Tanigawa,K., Kanlyakrit,W., and Hayashida,S. (1994b). The mechanism of binding of glucoamylase 1 from *Aspergillus-awamori* var. *kawachi* to cyclodextrins and raw starch. Biosci. Biotechnol. Biochem. *58*, 49-54.

Guillen,D., Santiago,M., Linares,L., Perez,R., Morlon,J., Ruiz,B., Sanchez,S., and Rodriguez-Sanoja,R. (2007). α-amylase starch binding domains: cooperative effects of binding to starch granules of multiple tandemly arranged domains. Appl. Environ. Microbiol. *73*, 3833-3837.

Harata,K., Haga,K., Nakamura,A., Aoyagi,M., and Yamane,K. (1996). X-ray structure of cyclodextrin glucanotransferase from alkalophilic *Bacillus sp.* 1011. Comparison of two independent molecules at 1.8 Å resolution. Acta Crystallogr. D Biol. Crystallogr. *52*, 1136-1145.

Hemker,M., Stratmann,A., Goeke,K., Schroder,W., Lenz,J., Piepersberg,W., and Pape,H. (2001). Identification, cloning, expression, and characterization of the extracellular acarbose-modifying glycosyltransferase, AcbD, from *Actinoplanes* sp. strain SE50. J. Bacteriol. *183*, 4484-4492.

Henrissat,B. and Davies,G. (1997). Structural and sequence-based classification of glycoside hydrolases. Curr. Opin. Struct. Biol. *7*, 637-644.

Howitt,C.A., Rahman,S., and Morell,M.K. (2006). Expression of bacterial starch-binding domains in *Arabidopsis* increases starch granule size. Funct. Plant Biol. *33*, 257-266.

Hua,Y.W., Chi,M.C., Huei-Fen,L., Kuo,L.Y., Ku,K.L., and Lin,L.L. (2005). Adsorption-elution purification of chimeric *Bacillus stearothermophilus* leucine aminopeptidase II with raw-starch-binding activity. World J. Microbiol. Biotechnol.*21*, 689-694.

Iefuji,H., Chino,M., Kato,M., and Iimura,Y. (1996). Raw-starch-digesting and thermostable α-amylase from the yeast *Cryptococcus* sp. S-2: purification, characterization, cloning and sequencing. Biochem. J. *318 (Pt 3)*, 989-996.

Jacks,A.J., Sorimachi,K., Le Gal-Coëffet,M.F., Williamson,G., Archer,D.B., and Williamson,M.P. (1995). ^{1}H and ^{15}N assignments and secondary structure of the starch-binding domain of glucoamylase from *Aspergillus niger*. Eur. J. Biochem. *233*, 568-578.

Jane,J., Kasemsuwan,T., Leas,S., Zobel,H., and Robyt,J.F. (1994). Anthology of starch granule morphology by scanning electron microscopy. Starch-Stärke *46*, 121-129.

Janecek,S. (2002). A motif of a microbial starch-binding domain found in human genethonin. Bioinformatics *18*, 1534-1537.

Janecek,S. and Sevcik,J. (1999). The evolution of starch-binding domain. FEBS Lett. *456*, 119-125.

Janecek,S., Svensson,B., and MacGregor,E.A. (2003). Relation between domain evolution, specificity, and taxonomy of the α-amylase family members containing a C-terminal starch-binding domain. Eur. J. Biochem. *270*, 635-645.

Ji,Q., Oomen,R.J.F., Vincken,J.P., Bolam,D.N., Gilbert,H.J., Suurs,L.C.J.M., and Visser,R.G.F. (2004). Reduction of starch granule size by expression of an engineered tandem starch-binding domain in potato plants. Plant Biotechnol. J. *2*, 251-260.

Ji,Q., Vincken,J.P., Suurs,L.C.J.M., and Visser,R.G.F. (2003). Microbial starch-binding domains as a tool for targeting proteins to granules during starch biosynthesis. Plant Mol. Biol. *51*, 789-801.

Jørgensen,A.D., Nohr,J., Kastrup,J.S., Gajhede,M., Sigurskjold,B.W., Sauer,J., Svergun,D.I., Svensson,B., and Vestergaard,B. (2008). Small angle X-ray studies reveal that *Aspergillus niger* glucoamylase has a defined extended conformation and can form dimers in solution. J. Biol. Chem. *283*, 14772-14780.

Juge,N., Nohr,J., Le Gal-Coëffet,M.F., Kramhoft,B., Furniss,C.S., Planchot,V., Archer,D.B., Williamson,G., and Svensson,B. (2006). The activity of barley α-amylase on starch granules is enhanced by fusion of a starch binding domain from *Aspergillus nige*r glucoamylase. Biochim. Biophys. Acta *1764*, 275-284.

Kerk,D., Conley,T.R., Rodriguez,F.A., Tran,H.T., Nimick,M., Muench,D.G., and Moorhead,G.B. (2006). A chloroplast-localized dual-specificity protein phosphatase in Arabidopsis contains a phylogenetically dispersed and ancient carbohydrate-binding domain, which binds the polysaccharide starch. Plant J. *46*, 400-413.

Klein,C., Hollender,J., Bender,H., and Schulz,G.E. (1992). Catalytic center of cyclodextrin glycosyltransferase derived from X-ray structure analysis combined with site-directed mutagenesis. Biochemistry *31*, 8740-8746.

Klein,C. and Schulz,G.E. (1991). Structure of cyclodextrin glycosyltransferase refined at 2.0 Å resolution. J. Mol. Biol. *217*, 737-750.

Knegtel,R.M., Wind,R.D., Rozeboom,H.J., Kalk,K.H., Buitelaar,R.M., Dijkhuizen,L., and Dijkstra,B.W. (1996). Crystal structure at 2.3 Å resolution and revised nucleotide sequence of the thermostable cyclodextrin glycosyltransferase from *Thermonanaerobacterium thermosulfurigenes* EM1. J. Mol. Biol. *256*, 611-622.

Knegtel,R.M.A., Strokopytov,B., Penninga,D., Faber,O.G., Rozeboom,H.J., Kalk,K.H., Dijkhuizen,L., and Dijkstra,B.W. (1995). Crystallographic studies of the interaction of cyclodextrin glycosyltransferase from *Bacillus circulans* strain 251 with natural substrates and products. J. Biol. Chem. *270*, 29256-29264.

Kötting,O., Pusch,K., Tiessen,A., Geigenberger,P., Steup,M., and Ritte,G. (2005). Identification of a novel enzyme required for starch metabolism in Arabidopsis leaves. The phosphoglucan, water dikinase. Plant Physiol. *137*, 242-252.

Lawson,C.L., van Montfort.R., Strokopytov,B., Rozeboom,H.J., Kalk,K.H., de Vries,G.E., Penninga,D., Dijkhuizen,L., and Dijkstra,B.W. (1994). Nucleotide sequence and X-ray structure of cyclodextrin glycosyltransferase from *Bacillus circulans* strain 251 in a maltose-dependent crystal form. J. Mol. Biol. *236*, 590-600.

Le Gal-Coëffet,M.F., Jacks,A.J., Sorimachi,K., Williamson,M.P., Williamson,G., and Archer,D.B. (1995). Expression in *Aspergillus niger* of the starch-binding domain of glucoamylase. Comparison with the proteolytically produced starch-binding domain. Eur. J. Biochem. *233*, 561-567.

Lerner,D.R. and Raikhel,N.V. (1992). The gene for stinging nettle lectin (*Urtica dioica* agglutinin) encodes both a lectin and a chitinase. J. Biol. Chem. *267*, 11085-11091.

Lin,L.L., Lo,H.F., Chi,M.C., and Ku,K.L. (2003). Functional expression of the raw starch-binding domain of *Bacillus sp* strain TS-23 α-amylase in recombinant *Escherichia coli*. Starch/Starke *55*, 197-202.

Liu,Y.N., Lai,Y.T., Chou,W.I., Chang,M.D., and Lyu,P.C. (2007). Solution structure of family 21 carbohydrate-binding module from *Rhizopus oryzae* glucoamylase. Biochem. J. *403*, 21-30.

Lo,H.F., Chiang,W.Y., Chi,M.C., Hu,H.Y., and Lin,L.L. (2004). Site-directed mutagenesis of the conserved threonine, tryptophan, and lysine residues in the starch-binding domain of *Bacillus sp* strain TS-23 α-amylase. Curr. Microbiol. *48*, 280-284.

Lo,H.F., Lin,L.L., Chiang,W.Y., Chie,M.C., Hsu,W.H., and Chang,C.T. (2002). Deletion analysis of the C-terminal region of the α-amylase of *Bacillus sp* strain TS-23. Arch. Microbiol. *178*, 115-123.

Lohi,H., Ianzano,L., Zhao,X.C., Chan,E.M., Turnbull,J., Scherer,S.W., Ackerley,C.A., and Minassian,B.A. (2005). Novel glycogen synthase kinase 3 and ubiquitination pathways in progressive myoclonus epilepsy. Hum. Mol. Genet. *14*, 2727-2736.

MacGregor,E.A., Janecek,S., and Svensson,B. (2001). Relationship of sequence and structure to specificity in the α-amylase family of enzymes. Biochim. Biophys. Acta *1546*, 1-20.

Machovic,M. and Janecek,S. (2006a). Starch-binding domains in the post-genome era. Cell Mol. Life Sci. *63*, 2710-2724.

Machovic,M. and Janecek,S. (2006b). The evolution of putative starch-binding domains. FEBS Lett. *580*, 6349-6356.

Machovic,M. and Janecek,S. (2008). Domain evolution in the GH13 pullulanase subfamily with focus on the carbohydrate-binding module family 48. Biologia *63*, 1057-1068.

Machovic,M., Svensson,B., MacGregor,E.A., and Janecek,S. (2005). A new clan of CBM families based on bioinformatics of starch-binding domains from families CBM20 and CBM21. FEBS J. *272*, 5497-5513.

Martin,C. and Smith,A.M. (1995). Starch biosynthesis. Plant Cell *7*, 971-985.

Mezaki,Y., Katsuya,Y., Kubota,M., and Matsuura,Y. (2001). Crystallization and structural analysis of intact maltotetraose-forming exo-amylase from *Pseudomonas stutzeri*. Biosci. Biotechnol. Biochem. *65*, 222-225.

Mikami,B., Adachi,M., Kage,T., Sarikaya,E., Nanmori,T., Shinke,R., and Utsumi,S. (1999). Structure of raw starch-digesting *Bacillus cereus* β-amylase complexed with maltose. Biochemistry *38*, 7050-7061.

Mikami,B., Iwamoto,H., Malle,D., Yoon,H.J., Mirkan-Sarikaya,E., Mezaki,Y., and Katsuya,Y. (2006). Crystal structure of pullulanase: evidence for parallel binding of oligosaccharides in the active site. J. Mol. Biol. *359*, 690-707.

Mikkelsen,R., Suszkiewicz,K., and Blennow,A. (2006). A novel type carbohydrate-binding module identified in α-glucan, water dikinases is specific for regulated plastidial starch metabolism. Biochemistry *45*, 4674-4682.

Minassian,B.A., Ianzano,L., Meloche,M., Andermann,E., Rouleau,G.A., Gado-Escueta,A.V., and Scherer,S.W. (2000). Mutation spectrum and predicted function of laforin in Lafora's progressive myoclonus epilepsy. Neurology *55*, 341-346.

Morris,V.J., Gunning,A.P., Faulds,C.B., Williamson,G., and Svensson,B. (2005). AFM images of complexes between amylose and *Aspergillus niger* glucoamylase mutants, native, and mutant starch binding domains: A model for the action of glucoamylase. Starch-Starke *57*, 1-7.

Nagasaka,Y., Kurosawa,K., Yokota,A., and Tomita,F. (1998). Purification and properties of the raw-starch-digesting glucoamylases from *Corticium rolfsii*. Appl. Microbiol. Biotechnol. *50*, 323-330.

Nanmori,T., Nagai,M., Shimizu,Y., Shinke,R., and Mikami,B. (1993). Cloning of the β-amylase gene from *Bacillus cereus* and characteristics of the primary structure of the enzyme. Appl. Environ. Microbiol. *59*, 623-627.

Nielsen,M.M., Seo,E.S., Bozonnet,S., Aghajari,N., Robert,X., Haser,R., and Svensson,B. (2008). Multi-site substrate binding and interplay in barley α-amylase 1. FEBS Lett. *582*, 2567-2571.

Niittyla,T., Comparot-Moss,S., Lue,W.L., Messerli,G., Trevisan,M., Seymour,M.D., Gatehouse,J.A., Villadsen,D., Smith,S.M., Chen,J., Zeeman,S.C., and Smith,A.M. (2006). Similar protein phosphatases control starch metabolism in plants and glycogen metabolism in mammals. J. Biol. Chem. *281*, 11815-11818.

Nitta,Y., Shirakawa,M., and Takasaki,Y. (1996). Kinetic study of the active site structure of β-amylase from *Bacillus cereus* var *mycoides*. Biosci. Biotechnol. Biochem. *60*, 823-827.

Norman,B.E., Pedersen,S., Bisgaard-Frantzen,H., and Borchert,T.K. (1997). The development of a new heat-stable α-amylase for calcium free starch liquefaction. Starch/Stärke *49*, 371-379.

Ohdan,K., Kuriki,T., Takata,H., Kaneko,H., and Okada,S. (2000a). Introduction of raw starch-binding domains into *Bacillus subtilis* α-amylase by fusion with the starch-binding domain of *Bacillus* cyclomaltodextrin glucanotransferase. Appl. Environ. Microbiol. *66*, 3058-3064.

Ohdan,K., Kuriki,T., Takata,H., and Okada,S. (2000b). Cloning of the cyclodextrin glucanotransferase gene from alkalophilic *Bacillus* sp A2-5a and analysis of the raw starch-binding domain. Appl. Microbiol. Biotechnol. *53*, 430-434.

Oyama,T., Kusunoki,M., Kishimoto,Y., Takasaki,Y., and Nitta,Y. (1999). Crystal structure of β-amylase from *Bacillus cereus* var. *mycoides* at 2.2 Å resolution. J. Biochem. *125*, 1120-1130.

Oyama,T., Miyake,H., Kusunoki,M., and Nitta,Y. (2003). Crystal structures of β-amylase from *Bacillus cereus* var. *mycoides* in complexes with substrate analogs and affinity-labeling reagents. J. Biochem. *133*, 467-474.

Page,R.D. (1996). TreeView: an application to display phylogenetic trees on personal computers. Comput. Appl. Biosci. *12*, 357-358.

Paldi,T., Levy,I., and Shoseyov,O. (2003). Glucoamylase starch-binding domain of *Aspergillus niger* B1: molecular cloning and functional characterization. Biochem. J. *372*, 905-910.

Palopoli,N., Busi,M.V., Fornasari,M.S., Gomez-Casati,D., Ugalde,R., and Parisi,G. (2006). Starch-synthase III family encodes a tandem of three starch-binding domains. Proteins *65*, 27-31.

Parsiegla,G., Schmidt,A.K., and Schulz,G.E. (1998). Substrate binding to a cyclodextrin glycosyltransferase and mutations increasing the γ-cyclodextrin production. Eur. J. Biochem. *255*, 710-717.

Pell,G., Williamson,M.P., Walters,C., Du,H., Gilbert,H.J., and Bolam,D.N. (2003). Importance of hydrophobic and polar residues in ligand binding in the family 15 carbohydrate-binding module from *Cellvibrio japonicus* Xyn10C. Biochemistry *42*, 9316-9323.

Penninga,D., Strokopytov,B., Rozeboom,H.J., Lawson,C.L., Dijkstra,B.W., Bergsma,J., and Dijkhuizen,L. (1995). Site-directed mutations in tyrosine 195 of cyclodextrin glycosyltransferase from *Bacillus circulans* strain 251 affect activity and product specificity. Biochemistry *34*, 3368-3376.

Penninga,D., van der Veen,B.A., Knegtel,R.M.A., van Hijum,S.A.F.T., Rozeboom,H.J., Kalk,K.H., Dijkstra,B.W., and Dijkhuizen,L. (1996). The raw starch binding domain of cyclodextrin glycosyltransferase from *Bacillus circulans* strain 251. J. Biol. Chem. *271*, 32777-32784.

Polekhina,G., Gupta,A., van Denderen,B.J., Feil,S.C., Kemp,B.E., Stapleton,D., and Parker,M.W. (2005). Structural basis for glycogen recognition by AMP-activated protein kinase. Structure *13*, 1453-1462.

Robert,X., Haser,R., Gottschalk,T.E., Ratajczak,F., Driguez,H., Svensson,B., and Aghajari,N. (2003). The structure of barley α-amylase isozyme 1 reveals a novel role of domain C in substrate recognition and binding: a pair of sugar tongs. Structure *11*, 973-984.

Robert,X., Haser,R., Mori,H., Svensson,B., and Aghajari,N. (2005). Oligosaccharide binding to barley α-amylase 1. J. Biol. Chem. *280*, 32968-32978.

Shinshi,H., Neuhas,J.M., Ryals,J., and Meins,F., Jr. (1990). Structure of a tobacco endochitinase gene: evidence that different chitinase genes can arise by transposition of sequences encoding a cysteine-rich domain. Plant Mol. Biol. *14*, 357-368.

Sigurskjold,B.W., Svensson,B., Williamson,G., and Driguez,H. (1994). Thermodynamics of ligand binding to the starch-binding domain of glucoamylase from *Aspergillus niger*. Eur. J. Biochem. *225*, 133-141.

Simpson,P.J., Jamieson,S.J., Abou Hachem,M., Karlsson,E.N., Gilbert,H.J., Holst,O., and Williamson,M.P. (2002). The solution structure of the CBM4-2 carbohydrate binding module from a thermostable *Rhodothermus marinus* xylanase. Biochemistry *41*, 5712-5719.

Smith,S.M., Fulton,D.C., Chia,T., Thorneycroft,D., Chapple,A., Dunstan,H., Hylton,C., Zeeman,S.C., and Smith,A.M. (2004). Diurnal changes in the transcriptome encoding enzymes of starch metabolism provide evidence for both transcriptional and posttranscriptional regulation of starch metabolism in arabidopsis leaves. Plant Physiol. *136*, 2687-2699.

Sorimachi,K., Jacks,A.J., Le Gal-Coëffet,M.F., Williamson,G., Archer,D.B., and Williamson,M.P. (1996). Solution structure of the granular starch binding domain of glucoamylase from *Aspergillus niger* by nuclear magnetic resonance spectroscopy. J. Mol. Biol. *259*, 970-987.

Sorimachi,K., Le Gal-Coëffet,M.F., Williamson,G., Archer,D.B., and Williamson,M.P. (1997). Solution structure of the granular starch binding domain of *Aspergillus niger* glucoamylase bound to β-cyclodextrin. Structure *5*, 647-661.

Southall,S.M., Simpson,P.J., Gilbert,H.J., Williamson,G., and Williamson,M.P. (1999). The starch-binding domain from glucoamylase disrupts the structure of starch. FEBS Lett. *447*, 58-60.

Steichen,J.M., Petty,R.V., and Sharkey,T.D. (2008). Domain characterization of a 4-α-glucanotransferase essential for maltose metabolism in photosynthetic leaves. J. Biol. Chem. *283*, 20797-20804.

Stoffer,B., Frandsen,T.P., Busk,P.K., Schneider,P., Svendsen,I., and Svensson,B. (1993). Production, purification and characterization of the catalytic domain of glucoamylase from *Aspergillus niger*. Biochem. J. *292*, 197-202.

Sugimoto,H., Nakaura,M., Kosuge,Y., Imai,K., Miyake,H., Karita,S., and Tanaka,A. (2007). Thermodynamic effects of disulfide bond on thermal unfolding of the starch-binding domain of *Aspergillus niger* glucoamylase. Biosci. Biotechnol. Biochem. *71*, 1535-1541.

Sumitani,J., Tottori,T., Kawaguchi,T., and Arai,M. (2000). New type of starch-binding domain: the direct repeat motif in the C-terminal region of *Bacillus* sp no. 195 α-amylase contributes to starch binding and raw starch degrading. Biochem. J. *350*, 477-484.

Svensson,B. (1988). Regional distant sequence homology between amylases, α-glucosidases and transglucanosylases. FEBS Lett. *230*, 72-76.

Svensson,B., Jespersen,H., Sierks,M.R., and MacGregor,E.A. (1989). Sequence homology between putative raw-starch binding domains from different starch-degrading enzymes. Biochem. J. *264*, 309-311.

Svensson,B., Larsen,K., Svendsen,I., and Boel,E. (1983). The complete amino acid sequence of glycoprotein, glucoamylase G1. from *Aspergillus niger*. Carlsberg Res. Commun. *48*, 529-544.

Svensson,B., Pedersen,T.G., Svendsen,I., Sakai,T., and Ottesen,M. (1982). Characterization of two forms of glucoamylase from *Aspergillus niger*. Carlsberg Res. Commun. *47*, 55-69.

Szabo,L., Jamal,S., Xie,H., Charnock,S.J., Bolam,D.N., Gilbert,H.J., and Davies,G.J. (2001). Structure of a family 15 carbohydrate-binding module in complex with xylopentaose. Evidence that xylan binds in an approximate 3-fold helical conformation. J. Biol. Chem. *276*, 49061-49065.

Takahashi,T., Kato,K., Ikegami,Y., and Irie,M. (1985). Different behavior towards raw starch of 3 forms of glucoamylase from a *Rhizopus* sp. J. Biochem. *98*, 663-671.

Takao,S., Sasaki,H., Kurosawa,K., Tanida,M., and Kamagata,Y. (1987). Production of a raw starch saccharifying enzyme by *Corticium rolfsii*. Agric. Biol. Chem. *50*, 1979-1987.

Tanaka,A., Karita,S., Kosuge,Y., Senoo,K., Obata,H., and Kitamoto,N. (1998). Thermal unfolding of the starch binding domain of *Aspergillus niger* glucoamylase. Biosci. Biotechnol. Biochem. *62*, 2127-2132.

Tomme,P., Creagh,A.L., Kilburn,D.G., and Haynes,C.A. (1996). Interaction of polysaccharides with the N-terminal cellulose-binding domain of *Cellulomonas fimi* CenC .1. Binding specificity and calorimetric analysis. Biochemistry *35*, 13885-13894.

Tomme,P., van,T.H., Pettersson,G., Van,D.J., Vandekerckhove,J., Knowles,J., Teeri,T., and Claeyssens,M. (1988). Studies of the cellulolytic system of *Trichoderma reesei* QM 9414. Analysis of domain function in two cellobiohydrolases by limited proteolysis. Eur. J. Biochem. *170*, 575-581.

Tung,J.Y., Chang,M.D., Chou,W.I., Liu,Y.Y., Yeh,Y.H., Chang,F.Y., Lin,S.C., Qiu,Z.L., and Sun,Y.J. (2008). Crystal structures of the starch-binding domain from *Rhizopus oryzae* glucoamylase reveal a polysaccharide-binding path. Biochem. J. *416*, 27-36.

Ueda,S. (1981). Fungal glucoamylases and raw starch digestion. Trends Plant Sci. 89-90.

Uitdehaag,J.C., Kalk,K.H., van,d., V, Dijkhuizen,L., and Dijkstra,B.W. (1999). The cyclization mechanism of cyclodextrin glycosyltransferase (CGTase) as revealed by a γ-cyclodextrin-CGTase complex at 1.8-Å resolution. J. Biol. Chem. *274*, 34868-34876.

Valdez,H.A., Busi,M.V., Wayllace,N.Z., Parisi,G., Ugalde,R.A., and Gomez-Casati,D.F. (2008). Role of the N-terminal starch-binding domains in the kinetic properties of starch synthase III from *Arabidopsis thaliana*. Biochemistry *47*, 3026-3032.

Viksø-Nielsen,A., Andersen,C., Hoff,T., and Pedersen S. (2006). Development of new α-amylases for raw starch hydrolysis. Biocatal. Biotransform. *24*, 121-127.

Wanderley,K.J., Torres,F.A., Moraes,L.M., and Ulhoa,C.J. (2004). Biochemical characterization of α-amylase from the yeast *Cryptococcus flavus*. FEMS Microbiol. Lett. *231*, 165-169.

Wang,J., Stuckey,J.A., Wishart,M.J., and Dixon,J.E. (2002). A unique carbohydrate binding domain targets the lafora disease phosphatase to glycogen. J. Biol. Chem. *277*, 2377-2380.

Wang,P., Singh,V., Xue,H., Johnston,D.B., Rausch,K.D., and Tumbleson,M.E. (2007). Comparison of Raw Starch Hydrolyzing Enzyme with Conventional Liquefaction and Saccharification Enzymes in Dry-Grind Corn Processing. Cereal Chemistry *84*, 10-14.

Williamson,G., Belshaw,N.J., and Williamson,M.P. (1992). *O*-Glycosylation in *Aspergillus* glucoamylase. Conformation and role in binding. Biochem. J. *282*, 423-428.

Williamson,M.P., Le Gal-Coëffet,M.F., Sorimachi,K., Furniss,C.S., Archer,D.B., and Williamson,G. (1997). Function of conserved tryptophans in the *Aspergillus niger* glucoamylase 1 starch binding domain. Biochemistry *36*, 7535-7539.

Wind,R.D., Uitdehaag,J.C., Buitelaar,R.M., Dijkstra,B.W., and Dijkhuizen,L. (1998). Engineering of cyclodextrin product specificity and pH optima of the thermostable cyclodextrin glycosyltransferase from *Thermoanaerobacterium thermosulfurigenes* EM1. J. Biol. Chem. *273*, 5771-5779.

Xie,H., Gilbert,H.J., Charnock,S.J., Davies,G.J., Williamson,M.P., Simpson,P.J., Raghothama,S., Fontes,C.M., Dias,F.M., Ferreira,L.M., and Bolam,D.N. (2001). *Clostridium thermocellum* Xyn10B carbohydrate-binding module 22-2: the role of conserved amino acids in ligand binding. Biochemistry *40*, 9167-9176.

Ye,Z., Miyake,H., Tatsumi,M., Nishimura,S., and Nitta,Y. (2004). Two additional carbohydrate-binding sites of β-amylase from *Bacillus cereus* var. *mycoides* are involved in hydrolysis and raw starch-binding. J. Biochem. *135*, 355-363.

Yoon,H.J., Hirata,A., Adachi,M., Sekine,A., Utsumi,S., and Mikami,B. (1999). Structure of the starch-binding domain of *Bacillus cereus* β-amylase. J. Microbiol. Biotechnol. *9*, 619-623.

Figure legends

Figure 1. Amino acid sequence alignment of CBM motifs from the families 20, 21, 48 and 53. The abbreviations of the source proteins are given in Table 1. The individual groups of SBD sequences are coloured as follows: CBM20 (green), CBM21 (red), CBM48 (blue) CBM53 (magenta). Eleven CBM20 *consensus* amino acid residues (Svensson et al., 1989) are highlighted as follows: two tryptophans of the starch binding site 1 in blue and substitutions to phenylalanine or tyrosine in turquoise; one tryptophan of the starch binding site 2 in red, substitutions to phenylalanine or tyrosine in magenta, and remaining eight residues in yellow. The tyrosines or phenylalanines forming CBM21 starch binding site 2 are shown in green. The additional well conserved phenylalanine in CBM20 and CBM48 is shaded in black. The invariantly conserved CBM21 lysine is shaded in grey. The alignment was done by Clustal-W at the European Bioinformatic Institute's server (http://www.ebi.ac.uk/) and then manually adjusted.

Figure 2. Evolutionary tree of CBM motifs from the families 20, 21, 48 and 53. The abbreviations of the source proteins are given in Table 1. The individual groups of SBD sequences are coloured as follows: CBM20 representatives - green; CBM21 representatives - red; CBM48 representatives - blue; CBM53 representatives - magenta. The tree is based on the alignment of complete CBM sequences (shown in Fig. 1) including the gaps. It was calculated as a Phylip tree-type using the neighbour-joining method (http://www.ebi.ac.uk/) and displayed by the program TreeView (Page, 1996).

Figure 3 A cartoon representation of the *Aspergillus niger* glucoamylase CBM20 structure (PDB code 1kul) showing binding site 1 (A) and binding site 2 (B). The cartoon is coated by a transparent molecular surface representation to give a topological perspective of the binding sites and the N- and the C-termini are colored in yellow and blue, respectively. The view in panel B is rotated about 180° along the long axis of the molecule and the flexible loop showing the largest conformational change upon ligand binding is colored in red. The aromatic ligand binding residues and selected residues implicated in ligand interaction at each site are shown as sticks.

Figure 4 Panel A depicts a surface represention of the GH13 cyclodextrin glucanotransferase from *Bacillus circulans* strain 251 in complex with maltose (PDB code 1cdg). This structure illustrates the close proximity of CBM20 binding site 2, represented by Y633 (yellow) and L600 (red) to the active site cleft and the catalytic nucleophile E257 (green). Three bound maltose molecules are shown as sticks at binding site 1, at binding site 2, and at the conserved site on the catalytic domain of CGTases (upper part of the molecule). The close up in panel B reveals architectural features of CBM20 binding site 2 with L600 protruding into the solvent and restricting the access to this site to single α-glucan chains. The structure was rendered using Pymol v0.99 software.

Tables

Table 1

Family	Abbreviation	Specificity	EC	Source	GenPept	Length
20	AMY_Aspka	α-amylase	3.2.1.1	*Aspergillus kawachii*	BAA22993	640
20	AMY_Bacsp	α-amylase	3.2.1.1	*Bacillus* sp. TS-23	AAA63900	613
20	AMY_Crcsp	α-amylase	3.2.1.1	*Cryptococcus* sp. S-2	BAA12010	631
20	AMY_Strgr	α-amylase	3.2.1.1	*Streptomyces griseus*	CAA40798	566
20	MGA_Bacst	maltogenic α-amylase	3.2.1.133	*Bacillus stearothermophilus*	AAA22233	719
20	M3H_Brasp	maltotriohydrolase	3.2.1.116	*Brachybacterium* sp. LB25	BAE94180	615
20	M4H_Psest	maltotetraohydrolase	3.2.1.60	*Pseudomonas stutzeri*	AAA25707	548
20	M5H_Psesp	maltopentaohydrolase	3.2.1.-	*Pseudomonas* sp. KO-8940	BAA01600	614
20	CGT_Bacci	cyclodextrin glucanotransferase	2.4.1.19	*Bacillus circulans* strain 251	CAA55023	713
20	CGT_Klepn	cyclodextrin glucanotransferase	2.4.1.19	*Klebsiella pneumoniae*	AAA25059	655
20	CGT_Thbth	cyclodextrin glucanotransferase	2.4.1.19	*Thermoanaerobacterium thermosulfurogenes*	AAB00845	710
20	CGT_Thcsp	cyclodextrin glucanotransferase	2.4.1.19	*Thermococcus* sp. B1001	BAA88217	739
20	CGT_Hafme	cyclodextrin glucanotransferase	2.4.1.19	*Haloferax mediterranei*	CAI46245	713
20	ACT_Actsp	acarviose transferase	2.4.1.19	*Actinoplanes* sp. 50/110	AAE37556	724
20	BMY_Bacce	β-amylase	3.2.1.2	*Bacillus cereus*	BAA34650	546
20	BMY_Bacme	β-amylase	3.2.1.2	*Bacillus megaterium*	CAB61483	545
20	BMY_Thbth	β-amylase	3.2.1.2	*Thermoanaerobacterium thermosulfurogenes*	AAA23204	515
20	GMY_Aspka	glucoamylase	3.2.1.3	*Aspergillus kawachii*	BAA00331	639
20	GMY_Aspni	glucoamylase	3.2.1.3	*Aspergillus niger*	AAB59296	640
20	GMY_Hypje	glucoamylase	3.2.1.3	*Hypocrea jecorina*	2VN7_A	599
20	GMY_Lened	glucoamylase	3.2.1.3	*Lentinula edodes*	AAF75523	571
20	GMY_Neucr	glucoamylase	3.2.1.3	*Neurospora crassa*	AAE15056	626
20	6AGT_Artgl	6-α-glucosyltransferase	2.4.1.-	*Arthrobacter globiformis*	BAD34980	965
20	4AGT_Soltu	4-α-glucanotransferase	2.4.1.25	*Solanum tuberosum* (potato)	AAR99599	948
20	GWD3_Arath	α-glucan, water dikinase	2.7.9.4	*Arabidopsis thaliana* (mouse-ear cress)	AY747068	1196
20	GEN_Homsa	genethonin-1	-	*Homo sapiens* (human)	AAC78827	358
20	LAF_Homsa	laforin	3.1.3.48/3.1.3.16	*Homo sapiens* (human)	AAG18377	331
20	GPD_Homsa	glycerophosphodiester phosphodiesterase	3.1.-.-	*Homo sapiens*	AAH27588	672
20	APU_Bacst	amylopullulanase	3.2.1.1/41	*Bacillus stearothermophilus*	AAG44799	2018
20	APU_Thbth	amylopullulanase	3.2.1.1/41	*Thermoanaerobacterium thermosulfurogenes*	AAB00841	1861
20	IGT_Bacci	isocyclomaltooligosaccharide glucanotransferase	2.4.1.-	*Bacillus circulans*	BAF37283	995
20	CE1_Pyrfu	carbohydrate esterase 1	-	*Pyrococcus furiosus*	AAL81232	404
20	CE1_Thcko	carbohydrate esterase 1	-	*Thermococcus kodakarensis*	BAD84711	449
21	AMY_Lipko	α-amylase	3.2.1.1	*Lipomyces kononenkoae*	AAC49622	624
21	AMY_Lipst	α -amylase	3.2.1.1	*Lipomyces starkeyi*	AAN75021	647
21	GMY_Arxad	glucoamylase	3.2.1.3	*Arxula adeninivorans*	CAA86997	624
21	GMY_Mucci	glucoamylase	3.2.1.3	*Mucor circinelloides*	AAN85206	609
21	GMY_Rhior	glucoamylase	3.2.1.3	*Rhizopus oryzae*	AAQ18643	604
21	PPRS_Cloac	protein phosphatase 1 regulatory subunit	-	*Clostridium acetobutylicum*	AAK76874	247
21	PPRS_Homsa	protein phosphatase 1 regulatory subunit	-	*Homo sapiens* (human brain)	AAH47502	299
21	PPRS_Sacce	protein phosphatase 1 regulatory subunit	-	*Saccharomyces cerevisiae*	CAA86906	538
48	AMPK1_Ratno	AMPK β1 subunit	-	*Rattus norvegicus* (rat)	AAH62008	270
48	AKIN1_Zeama	AKIN-β-γ-1 protein	-	*Zea mays* (maize)	AF276085	497
48	SEX4_Arath	starch excess 4 protein	3.1.3.48	*Arabidopsis thaliana* (mouse-ear cress)	AAN28817	379
48	SNF1_Orysa	SNF1-related regulatory subunit β1	-	*Oryza sativa* (rice)	ABF95644	295
48	GSs_Grija	glycogen synthase subunit	2.4.1.21	*Griffithsia japonica* (red alga)	AAM93999	201
48	GBE_Escco	glycogen branching enzyme	2.4.1.18	*Escherichia coli*	AAA23872	728
48	SBE_Horvu	starch branching enzyme	2.4.1.18	*Hordeum vulgare* (barley)	AAP72268	775
48	GBE_Sacce	glycogen branching enzyme	2.4.1.18	*Saccharomyces cerevisiae*	AAA34632	704
48	GBE_Homsa	glycogen branching enzyme	2.4.1.18	*Homo sapiens* (human)	AAA58642	702
48	MOTH_Brehe	maltooligosyl trehalohydrolase	3.2.1.141	*Brevibacterium helvolum*	AAB95369	589
48	MOTH_Sulso	maltooligosyl trehalohydrolase	3.2.1.141	*Sulfolobus solfataricus*	BAA11010	558
48	PUL_Klepn	pullulanase	3.2.1.41	*Klebsiella pneumoniae*	AAA25124	1102
48	PUL_Horvu	pullulanase (limit dextrinase)	3.2.1.41	*Hordeum vulgare* (barley)	AAD04189	904
48	ISO_Pseam	isoamylase	3.2.1.68	*Pseudomonas amyloderamosa*	AAA25854	771
48	ISO_Sulso	isoamylase	3.2.1.68	*Sulfolobus solfataricus*	AAK42273	718
48	ISO_Orysa	isoamylase	3.2.1.68	*Oryza sativa* (rice)	BAA29041	733
53	SS3a_Arath	starch-synthase III - copy 1	2.4.1.21	*Arabidopsis thaliana* (mouse-ear cress)	NP_172637	1025
53	SS3b_Arath	starch-synthase III - copy 2	2.4.1.21	*Arabidopsis thaliana* (mouse-ear cress)	NP_172637	1025
53	SS3c_Arath	starch-synthase III - copy 3	2.4.1.21	*Arabidopsis thaliana* (mouse-ear cress)	NP_172637	1025

Table 2. A summary of CBM20 three-dimensional structures

Specificity	Source	PDB code	Ligand	Form	Reference
Cyclodextrin glucano-transferase GH13	*Bacillus circulans 8*[b]	1CGT	free	Wt	(Klein and Schulz, 1991)
		1CGU	β-CD	Wt[c]	(Klein et al., 1992)
		5CGT	maltotriose	Wt[c]	(Parsiegla et al., 1998)
	Bacillus circulans strain 251[b]	1CXI	free	Wt	(Knegtel et al., 1995)
		1CDG	maltose	Wt	(Lawson et al., 1994)
		1CXH	Maltoheptaose	Wt	(Knegtel et al., 1995)
		1CXE	α-CD	Wt	(Knegtel et al., 1995)
		1CXK	γ-CD	Wt[c]	(Uitdehaag et al., 1999)
		1TCM	free	W616A	(Penninga et al., 1996)
	Bacillus sp. 1011[b]	1PAM	free	Wt	(Harata et al., 1996)
	Bacillus stearo-thermophilus no. 2	1CYG	free	Wt	not available
	Thermoanaerobacterium thermosulfurigenes EM1[b]	1CIU	free	Wt	(Knegtel et al., 1996)
		1A47	Maltohexaose inhibitor	Wt	(Wind et al., 1998)
Maltogenic α-amylase GH13	*Geobacillus stearo-thermophilus C599*	1QHP	Maltose	Wt	(Dauter et al., 1999)
β-amylase GH14	*Bacillus cereus*[b]	1B90	free	Wt	(Mikami et al., 1999)
		1B9Z	maltose	Wt	(Mikami et al., 1999)
		1CQY	free	SBD (418-516)	(Yoon et al., 1999)
Glucoamylase GH15	*Aspergillus niger*	1KUL	free	SBD (509-616)	(Sorimachi et al., 1996)
		1ACO	β-CD	SBD (509-616)	(Sorimachi et al., 1997)
	Hypocrea jecorina	2VN4	free	Wt	(Bott et al., 2008)
FLJ11085[a]	*Homo sapiens*	2Z0B	free	Wt	not available

[a]Putative glycerolphosphodiester phosphodiesterase. [b]For these enzymes representative structure entries are listed from the many available. [c]Mutation was made in the catalytic domain

Table 3: Interaction of CBM20 with soluble oligosaccharides

CBM20 (expression host)	Ligand and method	K_d (M)	ΔG (kJ mol^{-1})	ΔH (kJ mol^{-1})	- $T\Delta S$ (kJ mol^{-1})	Reference
GA-SBD (*A. niger*)	Maltose[a]	6.3×10^{-3}				(Belshaw and Williamson, 1993)
	Maltoheptaose[a]	3.8×10^{-4}				
	Maltododecaose[a]	10.5×10^{-6}				
	β-CD[a]	1.7×10^{-6}				(Belshaw and Williamson, 1991)
GA-SBD (*A. niger*)	β-CD[a]	9.0×10^{-6}				(Le Gal-Coëffet et al., 1995)
GA-SBD (*A. niger*)	β-CD[a]	14.4×10^{-6}	-27.6			(Williamson et al., 1997)
W590K		6.4×10^{-6}	-29.6			
W563K		28.0×10^{-6}	-26.0			
GA-SBD (*A. niger*)	Maltoheptaose[a]	23.0×10^{-6}	-26.5			(Giardina et al., 2001)
W590K		9.5×10^{-7}	-27.2			
W563K		16.8×10^{-6}	-34.4			
GA-SBD (*A. niger*)	β-CD[b]	1.8×10^{-6}	-27.2	-54.7	27.5	(Sigurskjold et al., 1994)
	Maltoheptaose[b]	2.0×10^{-4}	-21.3	-49.8	28.6	
GA-SBD (*E.coli*)	β-CD[b]	21.3×10^{-6}	-26.7	-58.5	31.8	(Sugimoto et al., 2007)
GA Full-length (*A. niger*)	β-CD[b]	18.9×10^{-6}	-27.1	-65.7	38.6	(Sigurskjold et al., 1994)
GA Full-length (*A. awamori*)	α-CD[a]	3.3×10^{-4}	-19.9			(Goto et al., 1994b)
	β-CD[a]	17.7×10^{-6}	-27.1			
	γ-CD[a]	1.4×10^{-4}	-21.8			
	Amylose DP 20[a]	99.5×10^{-6}	-22.8			
pH 7.0	β-CD[a]	19.2×10^{-6}	-26.9			
pH 9.0		23.9×10^{-6}	-26.3			
GA-SBD (*A. niger*) pH 5.5	β-CD[c]	7.5×10^{-6}				(Christiansen unpublished)
GWD3-SBD (*E. coli*) pH 5.5	β-CD[c]	3.8×10^{-4}				(Christiansen unpublished)

Affinity measurements were performed by [a]UV difference spectroscopy, [b]Isothermal titration calorimetry or [c]Surface plasmon resonance.

Table 4 Interaction of CBM20s with starch granules, cross linked amylose or soluble starch

CBM20 / full length protein	Ligand	K_d	B_{max}	Reference
GA-SBD[a]	Corn starch	19.6 µM	2.14 µmol/g	(Williamson et al., 1992)
GA-SBD[a]	Corn starch	3.2 µM	0.56 µmol/g	(Paldi et al., 2003)
GA-SBD[a]	Potato starch	3.3 µM	0.08 µmol/g	
β-amylase[b]	Corn starch	0.09 µM	16.1 µmol/g	(Ye et al., 2004)
W449F		2.14 µM	15.8 µmol/g	
W495F		2.08 µM	16.4 µmol/g	
W449F/W495F		2.34 µM	16.7 µmol/g	
GA-SBD fused to β-gal[a]	Corn starch	2.6 µM[f]	85 ± 11 mol/L resin[g]	(Dalmia et al., 1995)
GA-SBD fused to β-gal[a]	Cross linked amylose	0.2 µM[f]	3.5 ± 0.2 mol/L resin[g]	
CGTase-SBD fused to β-gal[c]	Corn starch	2.1 µM[f]	42 ± 1.6 mol/L resin[g]	
CGTase-SBD fused to β-gal[c]	Cross linked amylose	0.4 µM[f]	3.7 ± 0.2 mol/L resin[g]	
CGTase SBD[d]	Soluble starch	7.7 µM[f]	-	(Ji et al., 2004)
CGTase double SBD[d]		0.7 µM[f]	-	
CGTase[d]	Potato starch	0.79 mg/mL	1.00[h]	(Penninga et al., 1996)
W616A		1.56 mg/mL	0.63[h]	
W616A/W662A		2.36 mg/mL	0.49[h]	
Y633A		1.52 mg/mL	0.88[h]	
Barley α-amylase SBD[a]	Barley starch	0.13 mg/mL		(Juge et al., 2006)

CBM20 / full length protein	Ligand	Relative adsorption	Reference
α-amylase[e] Δ 608	Corn starch	62.7 %[j]	(Lo et al., 2002)
Δ 540		~16 %[j]	
Δ 514		< 2 %[j]	
Δ 455		0 %[j]	
CGTase	Corn starch	68.5 %[j]	(Ohdan et al., 2000b)
GA-SBD fused to β-gal[a]	Corn starch	18 mL/g	(Chen et al., 1991)
α-amylase-SBD[e]	Corn starch	11.5 mL/g	(Lin et al., 2003)

[a]Glucoamylase from *A. niger*, [b]β-amylase from *B. cereus* var. *mycoides,* [c]CGTase from *B. macerans* [d]CGTase from *B. circulans* strain 251, [e]α-amylase from *Bacillus* sp. strain TS-23, [f]Converted from K_a values, [g]B_{max} is expressed as mol/L resin, where 1 L resin is a suspension of 0.1 g corn starch or 0.01 g cross linked amylose, [h]B_{max} is the maximal fraction of protein bound to raw starch, [i]B_{max} is defined as a saturation type rate, [j]Adsorption rate is defined as [(B-A)/B] x 100, where A indicates the residual protein after adsorption and B represents the protein concentration in the original enzyme solution. Δ decribes a truncation of the full length protein.

Figure 1

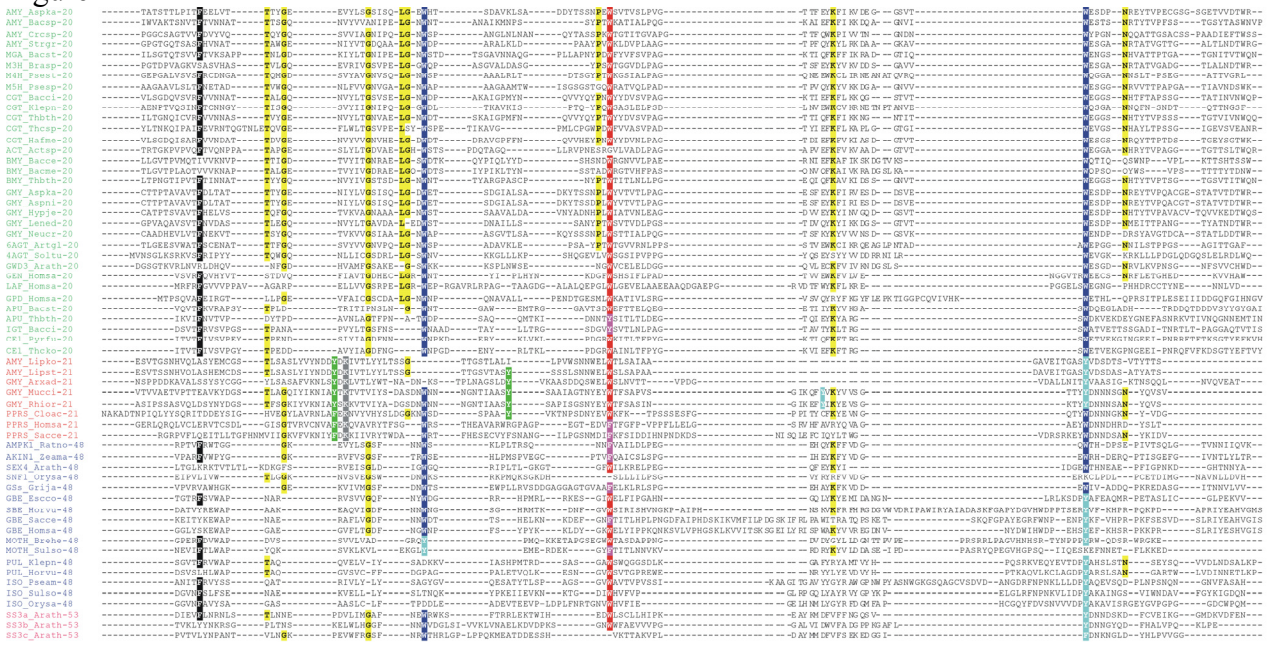


Figure 2

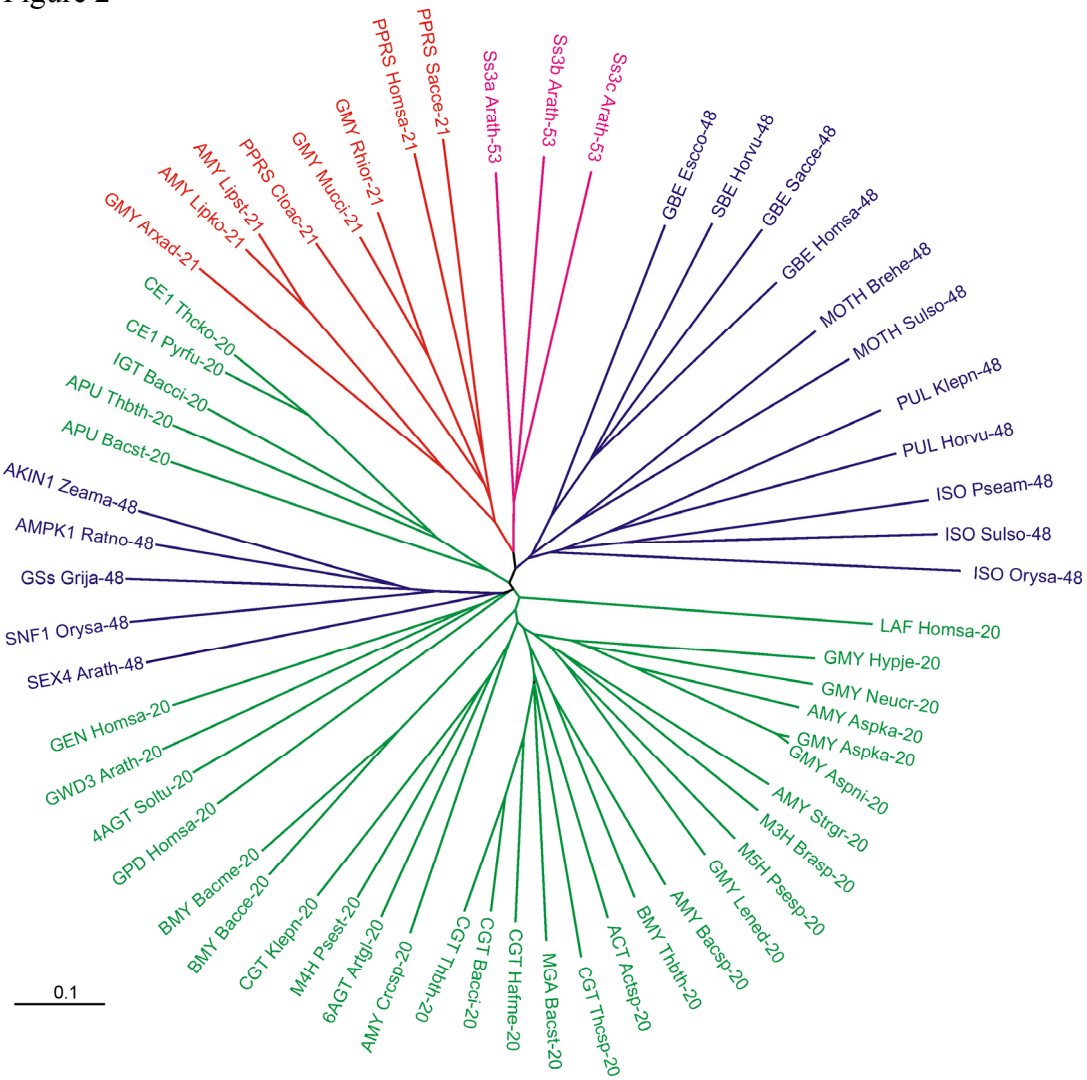

PPRS Homsa-21
PPRS Sacce-21
Ss3a Arath-53
Ss3b Arath-53
Ss3c Arath-53
GBE Escco-48
SBE Horvu-48
GBE Sacce-48
GBE Homsa-48
MOTH Brehe-48
MOTH Sulso-48
PUL Klepn-48
PUL Horvu-48
ISO Pseam-48
ISO Sulso-48
ISO Orysa-48
LAF Homsa-20
GMY Hypje-20
GMY Neucr-20
AMY Aspka-20
GMY Aspka-20
GMY Aspni-20
AMY Strgr-20
M3H Brasp-20
M5H Psesp-20
GMY Lened-20
AMY Bacsp-20
BMY Thbth-20
ACT Actsp-20
CGT Thcsp-20
MGA Bacst-20
CGT Hafme-20
CGT Bacci-20
CGT Thbth-20
AMY Crcsp-20
6AGT Artgl-20
M4H Psest-20
CGT Klepn-20
BMY Bacce-20
BMY Bacme-20
GPD Homsa-20
4AGT Soltu-20
GWD3 Arath-20
GEN Homsa-20
SEX4 Arath-48
SNF1 Orysa-48
GSs Grija-48
AMPK1 Ratno-48
AKIN1 Zeama-48
APU Bacst-20
APU Thbth-20
IGT Bacci-20
CE1 Pyrfu-20
CE1 Thcko-20
GMY Arxad-21
AMY Lipko-21
AMY Lipst-21
PPRS Cloac-21
GMY Mucci-21
GMY Rhior-21
0.1

Figure 3

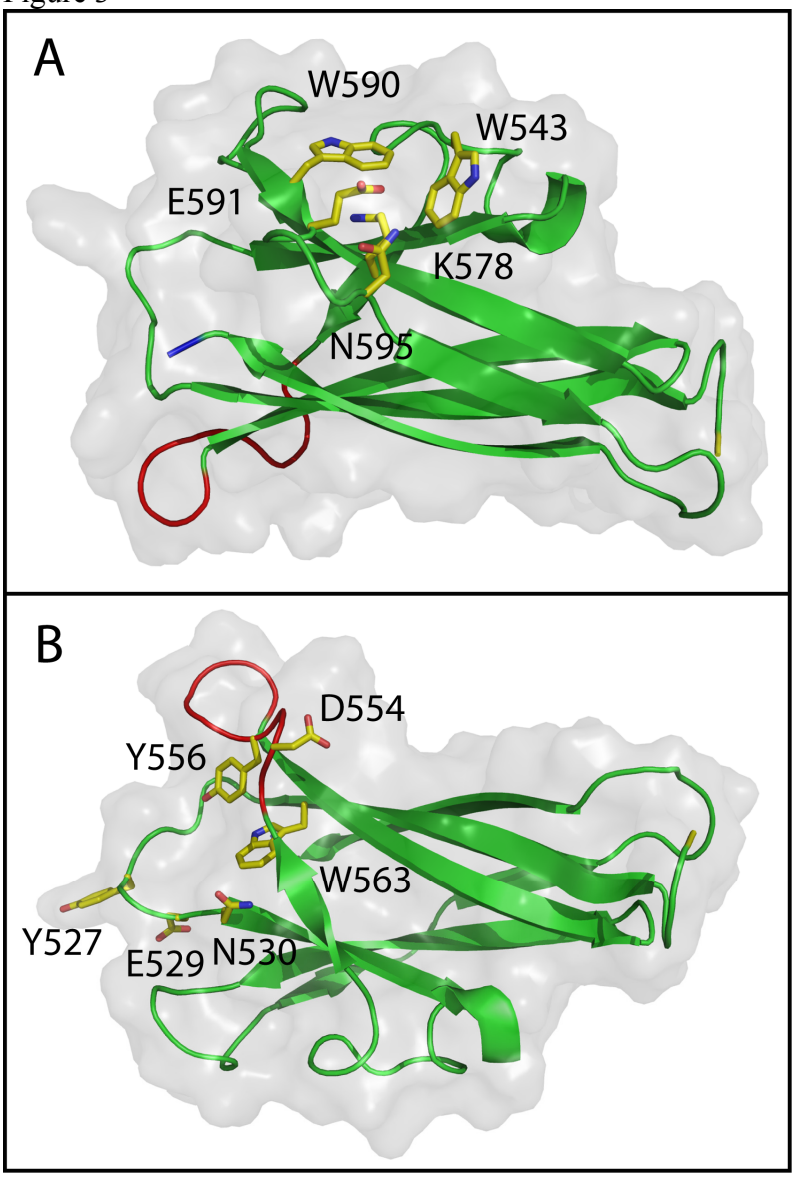
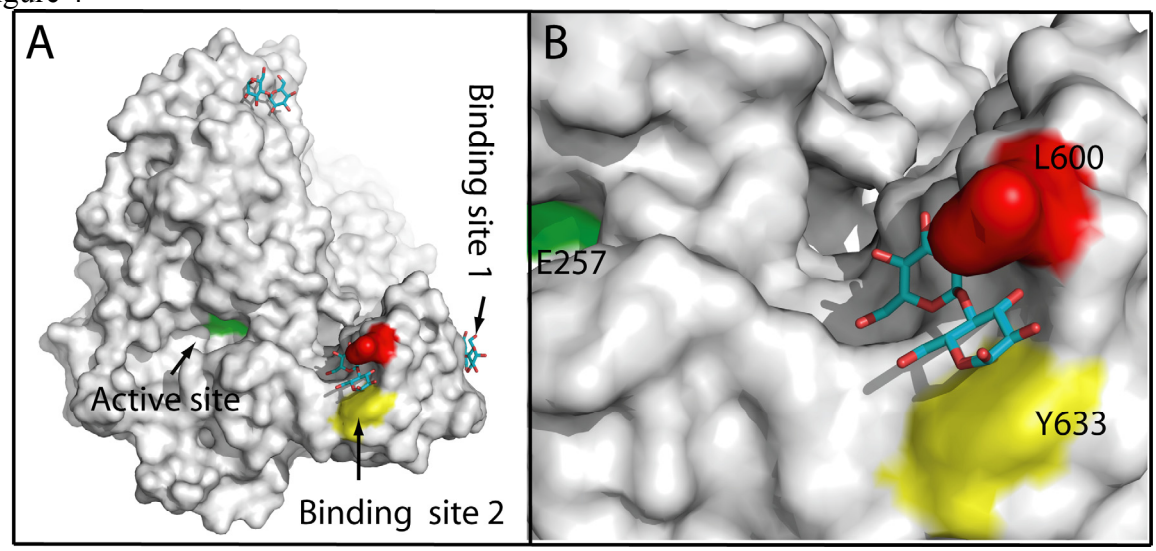
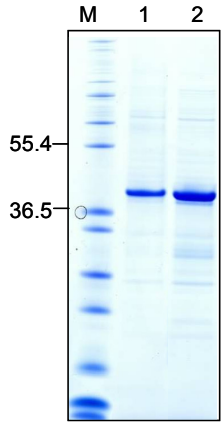


Figure 4

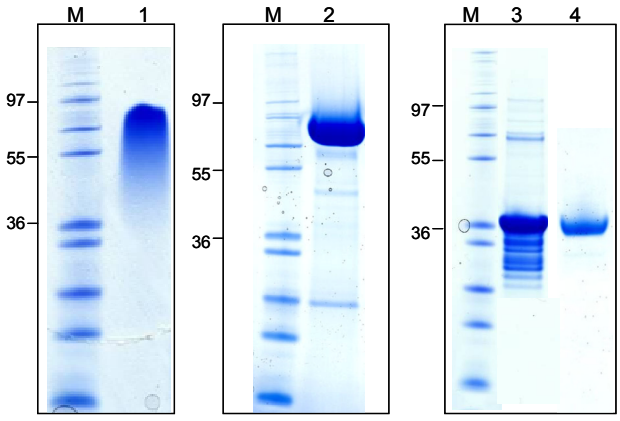


APPENDIX

Table 9.1: Primers used to amplify					
Protein	Organism	Vector	Sequence	Primer	Primer sequence
GWD3 SBD	*E. coli*	pDONR201/ and pDEST: 14,15 and 17	(68-170)	P1	F: GGGGACAAGTTTGTACAAAAAAGCAGGCT TCATGGATGGATCAGGAACGAAAGTGAGGT
				P2	R: GGGGACAAGTTTGTACAAGAAAGCTGGGT CCTAGGTTTCTCTAGTAGCATCCCAATGA
		pDONR201	F19A (68-170)	P3	F: GGTTAGATCATCAAGTTAAT**gcg**GGTGACC ATGTGGCTATGTTTGG
				P4	R: CCAAACATAGCCACATGGTCACC**cgc**ATTA ACTTGATGATCTAACC
		pDONR201	W35T (68-170)	P5	F: GGATCAGCTAAAGAGATTGGTTCA**ac**GAA AAAGAAATCGCCTTTGAATTGG
				P6	R: CCAATTCAAAGGCGATTTCTTTTTC**gt**TGAA CCAATCTCTTTAGCTGATCC
		pDONR201	W43T (68-170)	P7	F : GGAAAAAGAAATCGCCTTTGAAT**ac**GAGT GAGAATGGATGGGTTTG
				P8	R: CAAACCCATCCATTCTCACTC**gt**ATTCAAA GGCGATTTCTTTTTCC
		pDONR201	W75T (68-170)	P9	F: GAATGATGGTTCACTTTCA**ac**GGAATCTGG TGATAATCGTGTCC
				P10	R: GGACACGATTATCACCAGATTCC**gt**TGAAA GTGAACCATCATTC
		pET28a	(68-179 and 68-185)	P11	F: AACATG**CCATGG**CAGATGGATCAGGAAC GAAAGTG
			(68-179)	P12	R: AACCG**CTCGAGA**CCAACCTCCTGAGGCAAATC
			(68-185)	P13	R: AACCG**CTCGAGA**CCAACATCATCATCATTAC CAAC
	P. pastoris	pPICzα B	(68-170)	P14	F:ACGT**GGCCCAGCCGGCC**GATGGGTCAGGAAC GAG
				P15	R: GCC**TCTAGA**AAGGTTTCTCTAGTAGCATC
GA SBD	*E. coli*	pDONR201/ pDEST14-15	GA (509-616)	P16	F: GGGGACAAGTTTGTACAAAAAAGCAGGCTT CATGTGTACCACTCCCACCGCCGTGGCTG
				P17	R: GGGGACAAGTTTGTACAAGAAAGCTGGGTC CTACCGCCAGGTGTCAGTCACCGTC
	P. pastoris	pPICzα B	GA (509-616)	P18	F: ACGT**GGCCCAGCCGGCC** ATGTCGTTCCGA TCTCTA
				P19	R: GCC**TCTAGA**ACCGCCAGGTGTCAGTCA
LAF SBD	*E. coli*	pDONR201/ pDEST15	LAF-SBD (118-251)	P20	F: GGGGACAAGTTTGTACAAAAAAGCAGGCTT CATGCGCTTCCGCTTTGGGGTGGTGGTGCCA
				P21	R: GGGGACAAGTTTGTACAAGAAAGCTGGGTC CTACCTCAATCCAGTGTCCTATTGGGA
			LAF-SBD (138-251)	P22	F: GGGGACAAGTTTGTACAAAAAAGCAGGCTT CATGGAGCTGCTGGTGGTGGGGTCGCGGC
				P21	R: GGGGACAAGTTTGTACAAGAAAGCTGGGTC CTACCTCAATCCAGTGTCCTATTGGGA
GA	*P. pastoris*	GA-full lengh	GA (1-616)	P23	F: ACGT**GGCCCAGCCGGCC**ATGTCGTTCCGA TCTCTA
				P19	R: GCC**TCTAGA**ACCGCCAGGTGTCAGTCA
GA Without SBD	*P. pastoris*	GA without SBD	GA (1-508)	P23	F: ACGT**GGCCCAGCCGGCC**ATGTCGTTCCGATCTCTA
				P25	R: GCC**TCTAGA**AAGGAGGTTGATGACGTACTG
GA-GWD3 Fusion	*P. pastoris*	GA fused to GWD3-SBD	GA(1-508)+ GWD3 (68-170)	P23	F: ACGT**GGCCCAGCCGGCC**ATGTCGTTCCGATCTCTA
				P15	R: GCC**TCTAGA**AA GGTTTCTCTAGTAGCATC
				P25	Overlap primer F: CATCAACCTCCGATGGGTCAG
				P26	Overlap primer R: CTGACCCATCGGAGGTTGATG
Restriction site is shown in upper case letters and underlined. Mutant codon is shown in lower case letters and underlined					



Appendix Figure 13.1: SDS-PAGE analysis of purified *E. coli* pDEST15 GA and LAF.
Lane 1 GA-SBD after GSTrap purification, Lane 2: LAF-SBD after GSTrap purification.
Lane M: molecular weight standard



Appendix Figure 13.2: SDS-PAGE analysis of purified *GA, GA without SBD and GST*.
Lane 1: *P. pastoris* GA after HisTrap purification,
Lane 2: *P. pastoris* GA without SBD after HisTrap purification.
Lane 3: *E. coli* GST after GSTrap purification
Lane 4: *E. coli* GST from lane 3 followed by HisTrap purification
Lane M: molecular weight standard